DR 4.2: Multimodal Interaction with Basic Adaptivity and Personalization

Bernd Kiefer, Hans-Ulrich Krieger, Ivana Kruijff-Korbayová, Stefania Racioppa, Franziska Burger[†], Magdalena Kaiser[‡]

DFKI GmbH, TU Delft[†]*, Universität des Saarlandes*[‡]

⟨{kiefer,krieger,ivana.kruijff,stefania.racioppa}@dfki.de,
f.v.burger@tudelft.nl, s8makais@stud.uni-saarland.de⟩

Project, project Id:	EU H2020 PAL / PHC-643783
Project start date:	March 1 2015 (48 months)
Due date of deliverable:	M24
Actual submission date:	28 February 2017
Lead partner:	DFKI
Revision:	final
Dissemination level:	PU

The present report describes the work carried out in the second project year regarding *Natural Multimodal Interaction*. It summarises the Deliverable D4.2: “Natural Multimodal Interaction Basic Adaptivity and Personalisation”.

The formalism to implement dialogue strategies that was defined in the first year has been implemented to a great extend, and the functionality contained in the year 1 prototype has been implemented in the new formalism to prove the feasibility of the approach. On the natural language generation side, we implemented basic adaptivity and personalisation, using the data contained in the underlying knowledge source to adapt the conversational style, as well as the content, to the current user.

Two studies concerning personalised interaction were conducted. The first concerns the relation of the child’s disclosure depending on self-disclosure of the robotic agent, while the seconds develops strategies how to react to the child’s behaviour patterns in using the timeline.

Executive Summary The overall objective of WP4 is to support the goals set for a patient using the PAL system by developing the means to conduct verbal communication, and to analyse textual data and extract relevant information. The components implemented in this work package must support this communication in a way to foster sustainable long-term interactions between a robot (or its avatar) and a human. This requires user-adaptive communication, coupling of verbal and non-verbal communication and grounding communication in long-term memory. In the second year, we have extended the basic functionality for verbal interaction processing and developed strategies to conduct personalised interactions based on knowledge about the situation and a longer-term context.

This document describes the work done to reach deliverable D4.2, *Natural multimodal interaction with adaptivity and personalisation*, and the relevant milestones that are comprised in it.

We continued the implementation of the *dialogue management framework* developed in the first year and concluded the implementation of a first version of the compiler for the dialogue specifications and the run-time system. We are currently in the process to replace the prototype version of the dialogue management engine from year 1 by a module grounded in a declarative strategic dialogue rule base, which will improve the effectiveness of the dialogue management, and facilitate further adaptations that will lead to more broader and flexible dialogue.

The linguistic and semantic specifications in the ontology have been extended for the improved generation and interpretation of human language. In addition, we worked on an alternative representation of temporally changing information. The new representation will make it easier to access the data, and paves the way for a representation that contains uncertain or graded information, which is used ubiquitously in natural language.

For the *flexible and adaptive multimodal generation*, we extended our linguistic resources both for Italian and Dutch. We concluded the work on a free Dutch text-to-speech voice for the Mary TTS system, but found out that the quality was not sufficient for use in the PAL project. Since it is unclear how much resources would be necessary to obtain a voice that would fulfill the requirements, we decided to use a commercially available TTS system for Dutch instead.

For *role-adaptation between one-on-one and multi-party interaction*, we implemented a framework for multimodal fission for collaborative human-robot interaction. The frameworks automates the generation of deictic and referential natural language utterances for objects in the environment of a human-robot scenario, and synchronises these utterances with robot gaze and pointing gestures. This is also important for a multi-party scenario, where the robot gestures must be coherent with the current interaction sit-

uation.

The results of Year 2 are presented in 5 published peer-reviewed conference papers, two master theses, and a technical report.

1 The role of *Multimodal Natural Interaction* in PAL

WP4 focuses on the multimodal interaction around mHealth-Apps and additional conversational functionality in support of the high-level targets set in WP2 and actions selected in WP3. The challenge is to produce natural, flexible, personalised interactions that are sustainable in the long term as well as allow to extract data about the user. To achieve this, we are incorporating findings from the literature about what aspects are important for long term engagement.

The processing challenges for this work package are the robustness of input interpretation, flexibility *and* personal adaptation of the generated output, handling different situational contexts for both the physical and graphical embodiment of PAL, and allowing for interactions with one child alone or in the presence of an audience of multiple children. Additional challenges are posed by the need for extendable thematic and linguistic coverage.

The core functional component developed in WP4 is a multimodal dialogue system with a repertoire of multimodal dialogue acts (combining verbal and non-verbal means) modulated by affect. Generation as well as interpretation will use parameterised dialogue acts as an interface schema to other modules to abstract away from specific aspects of, e.g., natural language or emotion expression. Based on the high-level targets from WP2, action selection from WP3, the dialogue state (including the latest child's input and interaction history) as well as a long-term memory, the multimodal output generation module decides which act to activate ("what to express") and how to realise it multimodally in the given context ("how to express it").

In order to avoid repetitiveness in the long term, it is important to have flexible dialogue strategies and a rich repertoire of verbal and non-verbal expressions to allow for variation. The multimodal input processing module interprets verbal and non-verbal input. Interpretation is guided by information from the user model and the strategic planning (WP2 & 3) and provides information back to them. First, verbal input is needed for the dialog interaction itself as the dialogue's flow takes input from the child to progress. Second, an interpretation of the child's affective state is needed for engagement analysis used in WP2 to adapt the high-level goal self-management goals. Third, feedback is needed for WP3 as basis for adapting a child's preference model, and the long-term memory.

2 Tasks, objectives, results

The overall objective of WP4 is to develop the technologies for personalised multimodal natural interaction serving to actively foster long-term engagement with the robot and its avatar. Voluntary long-term use is required as a prerequisite for other system objectives. This encompasses natural language interpretation and multimodal generation, as well as dialogue management

2.1 Planned work

Our goals during Year 2 have been mainly to extend the interaction components to support the conducted experiments, and to improve the bond and the exchange between the user and the robotic agent.

The formalism for dialogue specification developed in the first year, which builds on a uniform knowledge representation to define and store information about and the structure of natural language and multimodal dialogue needed to be implemented to allow to build more flexible dialogue strategies in shorter time, and to improve modularization and reuse of existing dialogue modules. We have built a compiler for the declarative dialogue strategy rules, and a basic run-time system which can be used in various application contexts because it is agnostic to the underlying communication infrastructure.

Existing linguistic resources had to be extended to support more elaborate interactions with the user, which will in turn lead to deeper understanding, more flexible generation, and better adaptation to the user. During the experiments, interaction data with real users have been collected.

For upcoming experiments with multi-party interaction, we studied and implemented a framework for multi-modal fission that automates the generation of optimal deictic and referential language utterances and synchronises these with robot gaze and pointing gestures, which will also help to compute the right posture for the robot when there is more than one participant.

3 Actual work performed

The main work items, which will be detailed out in the next sections, were the following:

- We looked at different modules for Automatic Speech Recognition (ASR), trying to find out which could be applicable in PAL
- The implementation of the dialogue management formalism, as defined in Year 1, and the re-implementation of the existing dialogue functionality in the new formalism. This will allow to improve and adapt dialogue strategies much faster than before, and the specifications much more transparent and modular.
- An extended version of the ontology for the representation of interaction as well as personal user and health data, which allows more complex interactions covering a larger domain
- A novel representation for temporally changing data that also comprises a representation of uncertainty, which turns out to be important in the health data domain, but also to capture the vagueness in the semantics of many utterances in every day life
- Adapting existing modules and resources, e.g., the HFC reasoner, the existing grammars, etc., for use in PAL
- Studying self-disclosure of the artificial agent to increase the bond with the young users
- Studying opportunities for topic-driven and spontaneous interactions initiated by the artificial agent to make it appear more lively, and show that it is aware of the situation
- A framework for multimodal fission of deictic and referential utterances with robot gaze and pointing gestures in collaborative human-robot environments was implemented

3.1 Assessment of different ASR modules

We implemented a module to assess various automatic speech recognition (ASR) modules, which also includes a noise-robust voice activity detection (VAD) component. This allows to leave the microphone continuously open and send only audio chunks recognised as speech data to the speech recogniser. The cloud speech APIs of Google and Nuance are integrated, as well as the Nuance Recogniser 10.5. We are currently starting experiments to see if the results for child speech are sufficiently reliable to be used for, e.g., sentiment analysis, or feedback for the adaptation and personalisation modules.

We did some preliminary experiments with adult speakers, which still have to be conducted with children for a reliable comparison.

Unfortunately, Google has in the meantime shut down that part of their service that was used in our module, and we are on the lookout for further candidate services.

3.2 Dialogue Management Platform

We implemented the first version of a new framework for dialogue management and natural language processing, whose specifications we defined in the first year. It links dialogue processing tightly with knowledge representation in an ontology, and uses these representations to implement a long term memory for interactions, the information obtained by the user, and in the longer perspective maybe also for the personality of the robotic agent.

The new framework will facilitate the creation and extension of the dialogue management component because the formalism provides much more convenient and grounded ways to describe dialogue strategies. Furthermore, the formalism contains many shortcuts for the description and use of semantic entities, like overloaded operators for subsumption or type comparison, and convenient means for the access to the underlying long-term and interaction memory.

The existing PAL prototype has been rebuild using the declarative specifications. This could be achieved in very little time, which was partly due to the fact that the necessary functionality was already known and implemented, but also because the formalism allows to specify the knowledge needed for dialogue management very concisely. The resulting rule base is also much more readable because of its reduced size, modularity, and more declarative nature.

The use of the complex data structure representation in the Resource Description Framework (RDF) database is simplified by providing an implementation layer that shields the implementer from the underlying details, and provides standard programming language metaphors together with convenient abbreviations for the use in a natural language interaction environment.

During implementation, we explored several ways to express the procedural knowledge, about the dialogue, trying to find the best compromise between expressivity, ease of use and compactness. This lead to some changes in the specification, which in turn introduced a delay in the implementation. After some iterations, we are now confident to have found means to conveniently and compactly express the constraint knowledge needed in a dialogue system. Originally, our formalism was inspired by the *Information State – Update* paradigm which, as of today, is the generally accepted approach to dialogue systems, but we discovered that the formalism could also serve to fully implement a *Belief-Desire-Intent* agent.

This unforeseen dual view on the formalism is very interesting since it allows to fuse the language interaction functionality with the general behaviour of an artificial agent in a natural and uniform way.

3.3 Dialogue act semantics, ontology, knowledge base

We have continued our work on how to represent dialogue and interaction data in a RDF semantic database. One important aspect was the representation of uncertain data, which is used ubiquitously in language. This uncertainty is often expressed through specific verbs, adverbs, adjectives, or even phrases in natural language, e.g., *Tom suspects that he's laughed at in school.*, or *I thought I probably would catch a cold.*

For that, we now favour a different representation over the one used at the beginning of the project, where we extended the traditional triple representation by attaching the time a fact was entered to the database (the so-called transaction time). Now, in addition, we add a *confidence token* and use a set of modal operators which are directly linked to expressions in the language, to express the uncertainty and also be able to reason about the resulting (uncertain) facts [7], [3].

Because traditional ontology editors do not allow the specification of ontologies that use more than triples, we developed the ontology editor ×-*Protégé* that can cope with our extensions ([8]). It supports the definition of Cartesian types to represent n-ary relations and relation instances. These types are needed to specify facts (database entries) augmented by transaction time and confidence.

In addition, work on the base ontology of the project was continued optimising the structure to support the representation of learning goals and activities. Finding an appropriate representation for this data is important because it forms the basis for the assessment of the child's progress. This progress data is now also used to produce targeted feedback to the child in the form of multimodal utterances, to increase her/his motivation and the bonding with the agent.

3.4 Linguistic Resources

Currently, we are using two different approaches for the generation of verbal output. Firstly, we have a template generation engine which is called Content Planner which turns the application semantics, as described above, into a natural language utterance. Every output may consist of a single or more sentences, which can be concatenated or used alternatively. The advantage is that it is easy to add verbalisations for new inputs, even for non-expert users.

However, planning a natural verbal output using simple strings can become uncomfortable in some cases: e.g. in Italian the adjective inflection is given by the child (or the *activity* name) gender; the type of some prepositions depends on the following word (*conosci la risposta* ***alla prima*** *domanda* vs. ***a questa*** *domanda*). Again, some of the canned text variants affect the word order of the planned utterance. The whole verbalisation depends on the interaction type: apart from having to adapt the inflection (singular vs. plural), group interactions might require some special utterances which are unwanted/unusual in the single interactions.

Each syntactical variant requires different strings and consequently *different outputs* in the canned text, which makes maintaining the rule consistency more difficult.

On the other hand, *logical forms* (LF), like the semantics in the OpenCCG grammar, represent the words found and sentences using only their canonical form, while attributes like gender or number are provided as feature, not as string. As features can be parameterised more easily than strings, i.e. just using simple variables, substituting the output strings of the canned text with logical forms makes the rule maintenance more comfortable.

In the latest version, the possibility to combine LF and strings was implemented, so that the questions and answers from the Quiz game database can be fully integrated in the LF grammar. Using the OpenCCG surface realiser, we then can compute the text of the utterance from the defined LF (the linguistic semantics), which in turn is the input for the TTS system. The work on grammar adaptation is described in detail in annex A.3.1

3.5 Self-disclosure of the artificial agent

Self-Determination Theory posits that feeling connected to others can be an important factor in determining intrinsic motivation. According to Social Penetration Theory, such a connection can be created through the reciprocal disclosure of information about the self. Thus, a first exploratory feasibility study for using small talk in PAL, namely self-disclosure statements from the agent, to weld a bond between the diabetic child and the agent was conducted in Year 1. To this end, the agent was outfitted with a personality and a background story, which then informed the creation of a large number of self-disclosure statements at various levels of intimacy. The statements are stored in our extended RDF database with a valence label (positive, negative, neutral), a category label (school, sports, social, food), an intimacy label (from 0 to 3 with 0 being not intimate and 3 being very intimate), a language tag, an agent gesture, and a prompt sentence with which to encourage the child to respond with a disclosure of their own. We used the child selecting to open the diabetes diary as a triggering event for the agent disclosure in the experiment. After hearing the prompt, children were given the opportunity to respond or to opt out. The agent would then end the dialogue by thanking the child either for responding or simply for listening and the diary would open.

The small talk interaction was tested over the course of approximately two weeks at home by 11 children who had previously used the PAL application without this feature. It was found that the relative amount of disclosures that children made to the avatar was an indicator for the relatedness children felt towards the agent at the end of the study. Girls were significantly more likely to disclose and children preferred to reciprocate avatar disclosures of lower intimacy. No relationship was found between the intimacy level of avatar disclosures and child disclosures. Children did not add more content to the diary with the small talk interaction than in the prior evaluation of the application without this interaction, but their consistency in using the application did not decrease between the evaluations. It also became apparent from the interactions that children quickly became aware of the limitations of the interaction, i.e. non-responsiveness of the agent to their disclosure content. In future work, the interaction should thus be enhanced with intelligence.

3.6 Guided Feedback for more Life-Likeness and Bonding

We were working on identifying dialogue situations and appropriate content for giving feedback about the child's activities in the MyPal App. An example for this feedback is that the robot now praises the child when entering data into the timeline, like meals, glycemia values, or activities. The feedback is targeted and context-aware, e.g., values that are entered will change the type of feedback, like appraisal for "good" values, or understanding and motivation in cases were the values are not optimal, to avoid awkwardness on the child's side.

Also, the PAL agent will make suggestions to go to sections which are less populated, and compliment the child if the timeline or other activities are used on a regular basis, which is defined in the underlying application, as preference values in the database.

The next step will be to make the reactions adaptive to the child's personality, but that needs feedback on the behaviour of the child, either direct, by reacting to the agent's utterances, or indirect, by analysing the child's next actions.

3.7 Role-adaptation between one-on-one and multi-party interaction

Adaptation for multi-party interaction comes in many facets. In natural language dialogue, important aspects are turn-taking, choice of addressee (which can be one or many), and expressing this choice in the utterance in the right way, among others. In a human-robot environment, one important aspect is that the natural language utterances are accompanied by appropriate robot gestures that support the expressed content.

To this purpose, we have implemented a framework for multimodal fission human-robot interaction that automatically generates deictic or referential expressions to objects in the scene together with focused gaze and pointing gestures. A user study has shown that humans can resolve the deictic expressions much faster with the gestures than without. This work immediately carries over to multi-party scenarios, where the robot gaze should go to the right addressee, and the pointing gesture to some object that is referenced by speech.

4 Conclusions

The main work in year 2 was dedicated to the implementation of the dialogue management platform. It turned out that this was a much more complicated enterprise than expected, but through discussion with fellow colleagues, we also learned that its architecture comprises many architectural aspects from other fields, especially agent programming. We want to study these similarities and see if the framework could also be used (with extensions, or in combination with other frameworks) for the implementation of agents.

Furthermore, we supported the PAL agent with more means to express himself, always targeted towards the support for the young users, assisting them to accomplish their goals in diabetes management.

Overall, the WP-tasks provided the following major outcomes:

- Implementation of a dialogue management framework with prospects in agent programming and specification
- A logical representation of uncertainty in language expressions, enabling reasoning about these facts in a sound way
- Research on self-disclosure and its effect on child self-disclosure and bonding
- Implementation of feedback strategies using timeline data to encourage and support children doing the routine documentation work for their diabetes management
- Research on automatic multimodal generation for deictic expressions

With these outcomes, work package 4 achieved milestone 4.2, *Basic verbal and non-verbal adaptivity, basic social dialogue, techniques for the exploitation of knowledge base and memory, basic affective state recognition.* The new dialogue management platform supports all aspects of this milestone, while the research on self-disclosure and feedback targets the social aspects of multi-modal interaction, and the work on deictic expressions and uncertainty in language focuses on the adaptivity. The restructured ontology, together with targeted feedback rules, addresses the use of memory for long-term interactions.

Work on affective state recognition was mainly executed and presented in work package 3.

References

[1] Franziska Burger. Agents sharing secrets: Self-disclosure in long-term child-avatar interaction. Master's thesis, Radboud University Nijmegen, the Netherlands, 2016.

[2] Magdalena Kaiser. Multimodal fission with a focus on collaborative human-robot interaction *(in progress)*. Master's thesis, Universität des Saarlandes, Campus, 66123 Saarbrücken, 2017.

[3] Hans-Ulrich Krieger. Capturing graded knowledge and uncertainty in a modalized fragment of owl. In *Proceedings of the 8th International Conference on Agents and Artificial Intelligence. International Conference on Agents and Artificial Intelligence (ICAART-2016), February 24-26, Rome, Italy*. INSTICC, INSTICC, 2016. Conference Website: http://www.icaart.org.

[4] Hans-Ulrich Krieger, Rifca Peters, Bernd Kiefer, Michael A. van Bekkum, Frank Kaptein, and Mark A. Neerincx. The federated ontology of the pal project. interfacing ontologies and integrating time-dependent data. In *8th International Joint Conference on Knowledge Engineering and Ontology Development. International Conference on Knowledge Engineering and Ontology Development (KEOD-2016), November 9-11, Porto, Portugal*. INSTICC, SCITEPRESS, 11 2016.

[5] Hans-Ulrich Krieger and Stefan Schulz. A modal representation of graded medical statements. In *Annie Foret; Glyn Morell; Reinhard Muskens; Rainer Oswald; Sylvain Pogodalla: Formal Grammar 2015/2016*, volume 9804 of *Lecture Notes in Computer Science, LNCS*, pages 1–17. Springer-Verlag, Berlin Heidelberg, 2016.

[6] Stefania Racioppa. Ontology-based content planner grammars for PAL. Technical report, DFKI GmbH, 2016.

[7] Michael A. van Bekkum, Hans-Ulrich Krieger, Mark A. Neerincx, Frank Kaptein, Bernd Kiefer, Rifca Peters, and Stefania Racioppa. Ontology engineering for the design and implementation of personal pervasive lifestyle support. In *Proceedings of the 12th International Conference on Semantic Systems. ACM International Conference on Semantic Systems (SEMANTiCS-2016), The Linked Data Conference, September 12-15*. ACM ICP, 2016.

[8] Christian Willms, Hans-Ulrich Krieger, and Bernd Kiefer. X-protege: An ontology editor for defining cartesian types to represent n-ary relations. In *Joint Second Workshop on Language and Ontology & Terminology and*

Knowledge Structures. Workshop on Language and Ontology (LangOnto-2016), located at 10th edition of the Language Resources and Evaluation Conference, 2016, May 23-28, Portoroz, Slovenia. ELRA, ELDA, 5 2016.

A Annexes

A.1 Published Peer-Reviewed Papers

A.1.1 Ontology Engineering for the Design and Implementation of Personal Pervasive Lifestyle Support

Abstract The PAL project is developing: (1) an embodied conversational agent (robot and its avatar); (2) applications for child-agent activities that help children from 8 to 14 years old to acquire the required knowledge, skills and attitude for adequate diabetes self-management; and (3) dashboards for caregivers to enhance their supportive role for this self-management learning process. A common ontology is constructed to support normative behavior in a flexible way, to establish mutual understanding in the human-agent system, to integrate and utilize knowledge from the application and scientific domains, and to produce sensible human-agent dialogues. This paper presents the general vision, approach, and state of the art.

Relation to WP Representing the data relevant to a patient in a convenient form is important to refer to it in conversations, and possibly reason with them to give explanations for situations that the patient perceives as problematic, or appraise the progress a person makes. The temporal aspect is particularly important since it is indispensable for a history of events the virtual agent can allude to.

Availability Unrestricted. Included in the public version of this deliverable (annex B.1) [7].

A.1.2 Capturing Graded Knowledge and Uncertainty in a Modalized Fragment of OWL

Abstract Natural language statements uttered in diagnosis (e.g., in medicine), but more general in daily life are usually graded, i.e., are associated with a degree of uncertainty about the validity of an assessment and is often expressed through specific verbs, adverbs, or adjectives in natural language. In this paper, we look into a representation of such graded statements by presenting a simple non-standard modal logic which comes with a set of modal operators, directly associated with the words indicating the uncertainty and interpreted through confidence intervals in the model theory. We complement the model theory by a set of RDFS-/OWL 2 RL-like entailment (if-then) rules, acting on the syntactic representation of modalized statements. Our interest in such a formalization is related to the use of OWL as the de facto language in today's ontologies and its weakness to

represent and reason about assertional knowledge that is uncertain or that changes over time.

Relation to WP Effectively representing the content of verbal utterances containing uncertainty and to be able to reason about it is important because these utterances are ubiquitous in every day speech. Also, a virtual agent has to be able to generate verbal output based on uncertain data, and thus needs to know the correlation between the uncertainty and the verbal realization.

Availability Unrestricted. Included in the public version of this deliverable (annex B.2) [3].

A.1.3 ×-Protégé: An Ontology Editor for Defining Cartesian Types to Represent n-ary Relations.

Abstract Arbitrary n-ary relations ($n >= 1$) can, in principle, be realized through binary relations obtained by a reification process which introduces new individuals to which the additional arguments are linked via "accessor" properties. Modern ontologies which employ standards such as RDF and OWL have mostly obeyed this restriction, but have struggled with it nevertheless. In Krieger & Willms (2015), we have laid the foundations for a theory-agnostic extension of RDFS and OWL and have implemented in the last year an extension of Protege, called ×-Protégé, which supports the definition of Cartesian types to represent n-ary relations and relation instances. Not only do we keep the distinction between the domain and the range of an n-ary relation, but also introduce so-called extra arguments which can be seen as position-oriented unnamed annotation properties and which are accessible to entailment rules. As the direct representation of n-ary relations abolishes RDF triples, we have backed up ×-Protégé by the semantic repository and entailment engine HFC which supports tuples of arbitrary length. ×-Protégé is programmed in Java and is made available under the Mozilla Public License.

Relation to WP This ontology editor enhances the usability of the extended reasoner HFC that is employed in the PAL project as database layer. The representation of knowledge that was chosed uses a non-standard quintuple form that adds transaction time and uncertainty to the standard subject–predicate–object information. Standard editors are not able to effectively describe these, which implies the need for ×-Protégé.

Availability Unrestricted. Included in the public version of this deliverable (annex B.3) [8].

A.1.4 A Modal Representation of Graded Medical Statements

Abstract Medical natural language statements uttered by physicians are usually *graded*, i.e., are associated with a degree of uncertainty about the validity of a medical assessment. This uncertainty is often expressed through specific verbs, adverbs, or adjectives in natural language. In this paper, we look into a representation of such graded statements by presenting a simple non-standard *modal logic* which comes with a set of modal operators, directly associated with the words indicating the uncertainty and interpreted through confidence intervals in the model theory. Our interest in such a formalization is related to the use of OWL as the de facto standard in (medical) ontologies today and its weakness to represent and reason about assertional knowledge that is uncertain or that changes over time. The approach is not restricted to medical statements, but is applicable to other graded statements as well.

Relation to WP One important source of information for the PAL system consist in medical assessments in written text. To analyse these texts and make the results readily available in an abstract, but adequate form may greatly support the decision making modules in PAL.

Availability Unrestricted. Included in the public version of this deliverable (annex B.4) [5].

A.1.5 The Federated Ontology of the PAL Project. Interfacing Ontologies and Integrating Time-Dependent Data

Abstract This paper describes ongoing work carried out in the European project PAL which will support children in their diabetes self-management as well as assist health professionals and parents involved in the diabetes regimen of the child. Here, we will focus on the construction of the PAL ontology which has been assembled from several independently developed sub-ontologies and which are brought together by a set of hand-written interface axioms, expressed in OWL. We will describe in detail how the triple model of RDF has been extended towards transaction time in order to represent time-varying data. Examples of queries and rules involving temporal information will be presented as well. The approach is currently in use in diabetes camps.

Relation to WP Representing the data relevant to a patient in a convenient form is important to refer to it in conversations, and possibly reason with them to give explanations for situations that the patient perceives as problematic, or appraise the progress a person makes. The temporal aspect is particularly important since it is indispensable for a history of events the virtual agent can allude to.

Availability Unrestricted. Included in the public version of this deliverable (annex B.5) [4].

A.2 Master Theses

A.2.1 Franziska Burger (2017), "Agents Sharing Secrets: Self-Disclosure in Long-Term Child-Avatar Interaction"

Abstract A key challenge in developing companion agents for children is keeping them interested after novelty effects wear off. Self-Determination Theory posits that motivation is sustained if the human feels related to the agent. According to Social Penetration Theory, such a bond can be welded through the reciprocal disclosure of information about the self. As a result of these considerations, we developed a disclosure dialog module to study the self-disclosing behavior of children in response to that of a virtual agent. The module was integrated into a mobile application with avatar presence for diabetic children and subsequently used by 11 children in an exploratory field study over the course of approximately two weeks at home. It was found that the relative amount of disclosures that children made to the avatar was an indicator for the relatedness children felt towards the agent at the end of the study.

Relation to WP Having the child share information with the virtual agent is a key indicator of bonding, which according to all findings will keep the child motivated to interact with it over a longer period of time. This is a central point for the success of the PAL system.

Availability Unrestricted. Included in the public version of this deliverable (annex C.1).

A.2.2 Magdalena Kaiser (2017), "Multimodal Fission with a Focus on Collaborative Human-Robot Interaction"

Abstract The aim of my master thesis is to create an extendable, domain-independent framework for Multimodal Fission (abr. MMF) with a special focus on humanoid robots as execution environment. The term Fission describes the process of selecting a combination of channels which should be used to output a specific information to the user. This paper will give an overview of a first version of the MMF framework. How to categorize modalities and how to realize the modality selection process is investigated, as well as how to enable multimodal references. Since the focus of this thesis is on Human-Robot Interaction, modalities like Speech, Pointing and Gaze are implemented. However, other modalities can be added easily. The framework receives a simple predicate as input and outputs a multimodal representation of the information which can be presented by a robot.

Relation to WP The thesis relates to the work package in many aspects. It facilitates multi-modal generation of deictic and referential utterances, and shows that the support of a second modality in addition greatly improves the user's understanding. In addition, it paves a way for multi-user interaction by automatically generating robot movements and gestures towards the currently prominent member of a multi-party setup.

Availability Unrestricted (in progress). Résumé included in the public version of this deliverable (annex C.2).

A.3 Technical Reports

A.3.1 Stefania Racioppa (2016), "Ontology-based Content planner Grammars for PAL"

Abstract This technical report describes the work on the Content Planner grammars in PAL. In the past year, the main focus lied on the alignment of the Speech act definitions to the frames and instances defined in the PAL ontology. Besides this, the Content Planner and OpenCCG grammars were constantly improved in terms of coverage and parsing precision.

Relation to WP Directly relates to task T4.1.

Availability Unrestricted. Included in the public version of this deliverable (annex D.1).

B Publicly available papers

Ontology Engineering for the Design and Implementation of Personal Pervasive Lifestyle Support

Michael A. van Bekkum
TNO
Kampweg 5,
3769 DE, Soesterberg, Netherlands
michael.vanbekkum@tno.nl

Hans-Ulrich Krieger
DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken, Germany
krieger@dfki.de

Mark A. Neerincx
TNO
Kampweg 5,
3769 DE, Soesterberg, Netherlands
mark.neerincx@tno.nl

Frank Kaptein
Delft University of Technology
P.O. Box 5,
2600 AA, Delft, Netherlands
f.c.a.kaptein@tudelft.nl

Bernd Kiefer
DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken, Germany
bernd.kiefer@dfki.de

Rifca Peters
Delft University of Technology
P.O. Box 5,
2600 AA, Delft, Netherlands
r.m.peters@tudelft.nl

Stefania Racioppa
DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken, Germany
stefania.racioppa@dfki.de

ABSTRACT
The PAL project[1] is developing an *embodied conversational agent* (robot and its avatar), and applications for child-agent activities that help children from 8 to 14 years old to acquire the required *knowledge, skills, and attitude* for adequate diabetes self-management. Formal and informal caregivers can use the PAL system to enhance their supportive role for this self-management learning process. We are developing a **common ontology** (i) to support normative behavior in a flexible way, (ii) to establish mutual understanding in the human-agent system, (iii) to integrate and utilize knowledge from the application and scientific domains, and (iv) to produce sensible human-agent dialogues. The common ontology is constructed by relating and integrating partly existing separate ontologies that are specific to certain contexts or domains. This paper presents the general vision, approach, and state of the art.

CCS Concepts
Computing methodologies → Artificial intelligence → Knowledge representation and reasoning → Ontology engineering

Keywords
Ontology engineering; common ontology; embodied conversational agent; DIT++ standard; HFC inference engine; human-agent dialogue.

1. INTRODUCTION
In Europe, an increasing number of about 140,000 children (<14 year) have Type 1 Diabetes Mellitus (T1DM) [4]. The **PAL** project develops an *Embodied Conversational Agent* (**ECA**: robot and its avatar) and several applications for child-agent activities (e.g., playing a quiz and maintaining a timeline with the agent) that help these children to enhance their self-management .

PAL is part of a joint cognitive system in which humans and agents share information and learn to improve self-management. The required sharing of (evolving) knowledge has four important challenges:

1. To address the values and norms of both the caregivers and the caretakers in their different contexts (e.g., diabetes regimes, privacy).
2. To establish mutual understanding between the different human stakeholders of the PAL system, e.g., the end-users (children, caregivers), researchers and developers (e.g. academics, engineers).
3. To acquire, utilize, and deploy knowledge about child's self-management support.
4. To support natural and personalized interaction between the humans and PAL system agents.

In PAL, we are trying to meet these four challenges by developing a **common ontology** as an integrated part of the system development. The ontology addresses the aforementioned challenges by (1) serving as a knowledge basis for requirements analysis, (2) providing an unambiguous vocabulary and communication between stakeholders, (3) supporting system implementation of knowledge-based reasoning functionalities and (4) serving as a basis for interoperability in human-agent interaction. Engineering this ontology is a systematic, iterative, and incremental development process. Firstly, available ontologies and approaches are assessed on relevance and, possibly, adapted and integrated for our purposes (cf. Section 2). Secondly, relevant theories and models of the concerning scientific research fields are identified and formalized for adoption in the ontology (cf. Section 3). Thirdly, the ontology is implemented in an artefact or prototype, and subsequently, tested and refined (cf. Sections 4 and 5).

[1] **PAL**, *Personal Assistant for healthy Lifestyle*, is an European Horizon-2020 project; http://www.pal4u.eu

2. Engineering PAL Ontologies for Diabetes Self-Management Support

Because PAL covers a large domain of interest, we have developed separate ontology models as high-level building blocks for smaller, more specific areas of interest (frames). We have subsequently modeled each frame by either developing a new ontology or by selecting relevant, already existing models from (global) libraries that are similar in scope to the frame of interest.

The frames we have identified and modeled so far are among others (1) human/machine roles/actors involved in self-management, (2) task/goal/activity that includes self-management activities, tasks, associated goals, and results and the setting they take place in, (3) diabetes self-management activities and games, (4) issues related to medical examinations (e.g., lab values), and (5) dialogue management through a combination of dialogue acts and shallow semantic frames. A more elaborate PAL ontology will also include interaction and behavior models of robot and avatar, a model for privacy of information of self-management activities, and a model to cover the agreements and social contracts between patient and avatar/robot and a model for emotion and sentiment that covers the emotional responses of both robot and child to interaction as well as the general state of mind of the child.

As a modelling strategy, we have turned to existing (global) libraries to cover the various frames. Although our frames of interest are typically generic in nature, pre-existing models for these frames may differ (slightly) in scope and/or intention and may thus be a partial fit to the intended scope of PAL. Whereas e.g. the self-management part of diabetes is a relevant topic, the entire professional medical diagnosis and treatment model of diabetes may not be relevant here. We have therefore adapted these models whenever required by either extending them when concepts are missing or by selectively downsizing them when there are details/concepts in the model that are irrelevant to the scope of PAL. An example of reuse is displayed in the adoption of the well-known ontology for task world models [10] in the frame for tasks/goals. We have used this model at the core, but extended it with the Group concept, as a collection of Agents. At the same time, the notion of (external) Events triggering Tasks, has been discarded.

In the PAL project, dedicated studies of models in the scientific research areas concerned are also being conducted in order to adequately represent the frames of interest. For supporting the social processes that are involved in self-management learning, PAL models relationships in terms of familiarity or intimacy, liking, attitude, and benevolence [1]. Cognitive processes, the diabetes knowledge and corresponding learning goals have been explicitly modeled for purposes of monitoring and reasoning (aiming at personalized feedback by the ECA).

The affective process and state of a child are represented by a child ontology that allows the PAL system to estimate emotions experienced by the child resulting from activities proposed by the ECA. For example, the ECA can propose to play a quiz with the child, and predict joy as the emotional state of the child when the child did well during the game. This requires a complex affective state to be stored that contains all the affective information, be they, e.g., emotions (short, intense episodes) or moods (prolonged period of time). Emotions in this case need to be related to both child and the activity that had this emotion as a consequence. Moods need to contain a timestamp, indicating during which period it was measured. This representation makes it possible to find correlations between activities and affect over a prolonged period of time.

3. Extended Representation

The goal of the PAL ontology is norm-compliancy, shared understanding, interpretation, reasoning, and the generation of speech acts (e.g. verbal utterances). The ontology is based on a *uniform representation* of an application semantics that uses *dialogue acts* and *shallow semantic frames*, being represented by an extension of the RDF/OWL format (triples/binary relations) [9]. In addition, all user and other data that influence multimodal generation are specified in the ontology which facilitates access and combination of the different bits of information. We have extended existing processing components, e.g., the reasoning engine *HFC* [5] from DFKI and its database layer which makes information available to the interaction management and analysis.

One important part of the PAL ontology combines dialogue acts utilizing the DIT++ standard [2] and semantic frames, loosely based on thematic relations [3], used in today's frameworks VerbNet, VerbOcean, or FrameNet. Here is a simplified version of the combined representation that will be built for the sentence *Would you like to play a quiz*?

```
Offer[sender=MYSELF, addressee=ROBOT, ... ,
      frame=(Asking,agent=MYSELF,
             patient=ROBOT, theme=Quiz]]
```

Dialogue acts as well as semantic frames may contain further properties not depicted here; e.g., to represent the *continuation of dialogue acts* via `follows` or to model `indirect speech` through `refersTo`. Both properties map, again, to dialogue acts that have been introduced earlier in the conversation. The seemingly redundant specification of both dialogue act `Offer` and frame `Asking` is motivated by the fact that a *positive* answer to the question (= `AcceptOffer`) still refers to the `Asking` frame (*I'm accepting the offer you had asked for = yes*).

We have also defined a new way to marry the RDF-based triple representation with **transaction time** [6], as known from temporal databases [8]. This is possible because the inference engine HFC is based on *general tuples (n-ary relations)*, instead of restricting itself to triple-based representations. In the implementation of *HFC*, we employ 8-byte long integers (XSD datatype `long`) to encode *milli* or even *nano* seconds w.r.t. a fixed starting point (viz., Unix Epoch time). As a consequence, given a time point *t*, the next *smallest* or *successor* time point would then be *t+1*, thus our time is *discrete*. Like in *valid time*, the original approach to *transaction time* makes use of temporal intervals in order to represent the time during which a fact is stored in the database, even though the ending time is not known in advance. This is indicated by the wildcard **?** which will later be overwritten by the concrete ending time. We deviate here from the interval view by specifying both the *starting time* when an ABox statement is *entered* to an ontology, and, via a separate statement, the *ending time* when the statement is *invalidated*. For this, we exploit the propositional modals ⊤ and ⊥ (see [6]).

This idea is shown in the following picture for a binary relation P. We write P(c,d,b,e) to denote the row <c,d,b,e> in the database table ℙ for relation P.

Time	Database View	Ontology View
⋮	⋮	⋮
t_1	add: $\mathtt{P(c,d,}t_1\mathtt{,?)}$	add: $\top$P(c,d)@t_1
⋮	⋮	⋮
t_2	overwrite: $\mathtt{P(c,d,}t_1\mathtt{,}t_2\mathtt{)}$	——
t_2+1	——	add: $\bot$P(c,d)@t_2+1
⋮	⋮	⋮

Figure 1: Tuple representation with transaction time

As we see from this picture, the invalidation in the ontology happens at t_2+1, whereas $[t_1, t_2]$ specifies the transaction time in the database. Clearly, the same transaction time interval for P(c,d) in the ontology can be derived from the two statements ⊤P(c,d)@ t_1 and ⊥P(c,d)@ t_2+1, assuming that there does not exist a ⊥P(c,d)@t, such that $t_1 \leq t \leq t_2$ (we can effectively query for this by employing an ValidInBetween test in our SPARQL-like queries). Extending ontologies by transaction time the way we proceed here gives us a means to easily encode *time series data*, i.e., allows us to record the *history* of data that changes over time. Time-stamped data such as x ⊤ P(c,d)@t is represented in *HFC* by a quintuple: `true c P d t`.

4. Application and Implementation

The implementation of the ontology models is done in a PAL system that consists of several modules. A *dialogue manager* is, e.g., responsible for conversations between child and avatar/robot, an *action-selection module* decides what the best actions are at a particular moment (e.g., when playing a game), while a *child model module* is able to reason on the (emotional) state of the child.

This modular setup of the system requires clearly defined knowledge representations for each of the modules: the set of PAL ontologies for diabetes self-management support. To this end, the system has all individual ontologies defined in the extended *HFC* reasoner and has connected semantically-equivalent concepts, properties, and individuals through OWL interface axioms, utilizing standard constructors, such as `rdfs:subClassOf` or `owl:equivalentProperty`, or by posing domain and range restrictions (e.g., `rdfs:domain`).

The ontology engineering in PAL is meant to be an iterative and incremental process, with continuous refinement and extension of the involved ontologies. The models are adjusted according to new insights and continue to be aligned with sources of information in the entire project. The development of the ontology thus runs in parallel with the system design it is supposed to support and the modular approach allows for testing and refining these incrementally.

The PAL system is currently used in hospitals, diabetes camps, and at home. The analysis of the first recorded data sets for children and caregivers that have used the PAL system from a few days to four weeks in Italy and the Netherlands, is underway. Based on the ontological concepts, we will be able to identify meaningful patterns in the data that will be used to improve the intelligence of PAL, both in the knowledge base with refined ontologies, as well as its associated reasoning mechanisms. E.g., for the provisions of personalized feedback based on the identified user profiles, addressing cultural differences [7].

5. Discussion

This paper presents the development of a common ontology that underpins the design and implementation of an ECA for children, that help these children to acquire the required knowledge, skills, and attitude for adequate diabetes self-management. This ontology is used to establish mutual understanding in the human-agent system, to integrate and utilize knowledge from the application and scientific domains, and to produce natural human-agent dialogues. A set of interconnected ontologies ("frames") have been constructed, each consisting of general concepts and their relations: (1) roles and actors, (2) task, goal, activity and context, (3) diabetes self-management, and (4) dialogue management. We developed the first version of an ontology which species the data structures to be used by the dialogue specifications, dialogue history, and information state, and adapted our reasoning components, so that this knowledge source can be used efficiently once the formalism specification is fully implemented. The current version of the ontology is available at the PAL ontology website (http://www.dfki.de/lt/onto/pal/).

6. Acknowledgements

The research described in this paper has been funded by the Horizon 2020 Framework Programme of the European Union within the project PAL (Personal Assistant for healthy Lifestyle) under Grant agreement no. 643783.

7. REFERENCES

[1] Altman, I. and Taylor, D. Social penetration theory. Holt, Rinehart & Mnston, New York. (1973).

[2] Bunt, H., Alexandersson, J., Choe, J.-W., Fang, A.C., Hasida, K., Petukhova, V., Popescu-Belis, A., and Traum, D. ISO 24617-2: A semantically-based standard for dialogue annotation. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*. (2012).

[3] Fillmore, C.J. (1977). The Case for Case Reopened. In *Grammatical Relations*. Syntax & Semantics. Academic Press.

[4] D. Freeborn, T. Dyches, S. O. Roper, and B. Mandleco. Identifying challenges of living with type 1 diabetes: child and youth perspectives. *Journal of clinical nursing, vol. 22, no. 13-14*. (2013), 1890–1898.

[5] Krieger, H.-U. An Efficient Implementation of Equivalence Relations in OWL via Rule and Query Rewriting. *Proceedings of the 7th International Conference on Semantic Computing (ICSC).* (2013).

[6] Krieger, H.-U. Integrating Graded Knowledge and Temporal Change in a Modal Fragment of OWL. In *Agents and Artificial Intelligence, edited by Jaap van den Herik & Joaquim Filipe. Lecture Notes in Computer Science.* Springer. To appear. (2016).

[7] Neerincx, A., Sacchitelli, F., Kaptein, R., van der Pal, S., Oleari, E., and Neerincx, M. A. (2016). Child's Culture-related Experiences with a Social Robot at Diabetes Camps. *Eleventh ACM/IEEE International Conference on Human Robot Interaction* IEEE Press. (2016), 485-486.

[8] Snodgrass, R.T. Developing Time-Oriented Database Applications in SQL. Morgan Kaufmann, San Francisco, CA. (2000).

[9] ter Horst, H. J. Completeness, decidability and complexity of entailment for RDF Schema and a semantic extension involving the OWL vocabulary. Journal of Web Semantics, 3 (2005), 79–115.

[10] Welie, M. Van, Veer, G. C. Van Der, & Eliëns, A. An ontology for task world models. *Design, Specification and Verification of Interactive Systems '98*. (1998), 57–70.

Capturing Graded Knowledge and Uncertainty in a Modalized Fragment of OWL

Hans-Ulrich Krieger
German Reasearch Center for Artificial Intelligence (DFKI), Saarbrücken, Germany
krieger@dfki.de

Keywords: Knowledge Representation and Reasoning; Uncertainty in AI; Description Logics & OWL; Ontologies; Graded Natural Language Statements; Representation of Controlled Natural Language.

Abstract: Natural language statements uttered in diagnosis (e.g., in medicine), but more general in daily life are usually *graded*, i.e., are associated with a degree of uncertainty about the validity of an assessment and is often expressed through specific *verbs*, *adverbs*, or *adjectives* in natural language. In this paper, we look into a *representation* of such graded statements by presenting a simple non-standard modal logic which comes with a set of modal operators, directly associated with the words indicating the uncertainty and interpreted through confidence intervals in the model theory. We complement the model theory by a set of RDFS-/OWL 2 RL-like entailment (*if-then*) rules, acting on the syntactic representation of modalized statements. Our interest in such a formalization is related to the use of OWL as the *de facto* language in today's ontologies and its weakness to represent and reason about assertional knowledge that is *uncertain* or that changes over time.

1 INTRODUCTION

Medical natural language statements uttered by physicians or other health professionals and found in medical examination letters are usually *graded*, i.e., are associated with a degree of uncertainty about the validity of a medical assessment. This uncertainty is often expressed through specific *verbs, adverbs, adjectives*, or even *phrases* in natural language which we will call gradation words ($\approx$ linguistic hedges); e.g., *Dr. X suspects that Y suffers from Hepatitis* or *The patient probably has Hepatitis* or *(The diagnosis of) Hepatitis is confirmed*.

In this paper, we look into a representation of such graded statements by presenting a simple non-standard modal logic which comes with a small set of *partially-ordered modal operators*, directly associated with the words indicating the uncertainty and interpreted through *confidence intervals* in the model theory. The work presented here addresses modalized propositional formulae in negation normal form which can be seen as a canonical representation of natural language sentences of the above form (a kind of a *controlled natural language*).

Our interest in such a formalization is related to the use of OWL in our projects as the *de facto standard* for (medical) ontologies today (to represent structural/terminological knowledge) and its *weakness* to represent and reason about assertional knowledge that is uncertain (Schulz et al., 2014) or that changes over time (Krieger, 2012). There are two principled ways to address such a restriction: *either* by sticking with the existing formalism (viz., OWL) and trying to find an encoding that still enables some useful forms of reasoning (Schulz et al., 2014); *or* by deviating from a defined standard in order to arrive, at best, at an easier, intuitive, and less error-prone representation (Krieger, 2012).

Here, we follow the latter avenue, but employ and extend the standard entailment rules from (Hayes, 2004; ter Horst, 2005; Motik et al., 2012) for positive binary relation instances in RDFS and OWL towards modalized *n*-ary relation instances, including negation. These entailment rules talk about, e.g., subsumption, class membership, or transitivity, and have been found useful in many applications. The proposed solution has been implemented for the binary relation case (extended triples, quads) in *HFC* (Krieger, 2013), a forward chaining engine that builds Herbrand models which are compatible with the open-world view underlying OWL.

Our approach is clearly not restricted to medical statements, but is applicable to graded statements in general, e.g., in technical diagnosis (*the engine is probably overheated*) or in everyday conversation (*I'm pretty sure that Joe has signed a contract with*

Foo Inc.), involving *trust* (*I'm not an expert, but ...*) which can be seen as the common case (contrary to true *universal* statements).

2 OWL VS. MODALIZED REPRESENTATION

We note here that the names of our *initial* modal operators were inspired by the *qualitative information parts* of diagnostic statements from (Schulz et al., 2014) as shown in Figure 1.

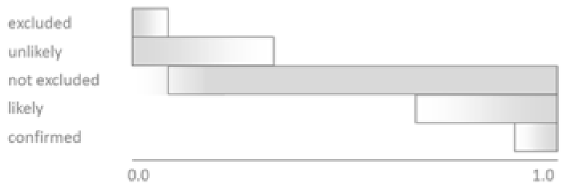


Figure 1: Schematic mappings of the qualitative information parts *excluded* (E), *unlikely* (U), *not excluded* (N), *likely* (L), and *confirmed* (C) to *confidence intervals*. Picture taken from (Schulz et al., 2014).

These qualitative parts were used in medical statements about, e.g., liver inflammation with varying levels of detail (Schulz et al., 2014) in order to infer, e.g., **if** *Hepatitis is confirmed* **then** *Hepatitis is likely* but **not** *Hepatitis is unlikely*. And **if** *Viral Hepatitis B is confirmed*, **then** both *Viral Hepatitis is confirmed* **and** *Hepatitis is confirmed* (generalization). Things "turn around" when we look at the adjectival modifiers *excluded* and *unlikely*: **if** *Hepatitis is excluded* **then** *Hepatitis is unlikely*, but **not** *Hepatitis is not excluded*. Furthermore, **if** *Hepatitis is excluded*, **then** both *Viral Hepatitis is excluded* **and** *Viral Hepatitis B is excluded* (specialization).

(Schulz et al., 2014) consider five OWL encodings, from which only two were able to fully reproduce the *plausible* inferences for the above Hepatitis use case. The encodings in (Schulz et al., 2014) were quite *cumbersome* as the primary interest was to stay within the limits of the underlying calculus. Besides coming up with complex encodings, only minor forms of reasoning were possible, viz., subsumption reasoning. Furthermore, each combination of disease and qualitative information part required a *new* OWL class definition/new class name, and there exist a lot of them!

These disadvantages are a result of two conscious decisions: OWL only provides unary and binary relations (concepts and roles) and comes up with a (mostly) fixed set of entailment/tableaux rules.

In our approach, however, the *qualitative information parts* from Figure 1 are first class citizens of the object language (the modal operators) and *diagnostic statements* from the Hepatitis use case are expressed through the binary property $\mathsf{suffersForm}$ between p (patients, people) and d (diseases, diagnoses). The plausible inferences are then simply a *byproduct* of the *instantiation* of the entailment rule schemas (G) from Section 5.1, and (S1) and (S0) from Section 5.2 for property $\mathsf{suffersForm}$ (the rule variables are universally quantified; $\top$ = *universal truth*; C = *confirmed*; L = *likely*), e.g.,

(S1) $\mathsf{ViralHepatitisB} \sqsubseteq \mathsf{ViralHepatitis} \wedge \top\mathsf{ViralHepatitisB}(d) \rightarrow \top\mathsf{ViralHepatitis}(d)$

(G) $C\mathsf{suffersFrom}(p,d) \rightarrow L\mathsf{suffersFrom}(p,d)$

Two things are worth mentioning here. *Firstly*, not only OWL properties can be graded, such as $C\mathsf{suffersFrom}(p,d)$ (= *it is confirmed that p suffers from d*), but also class membership, e.g., $C\mathsf{ViralHepatitisB}(d)$ (= *it is confirmed that d is of type Viral Hepatitis B*). As the original OWL example from (Schulz et al., 2014) can not make use of any modals, we employ the special modal $\top$ here: $\top\mathsf{ViralHepatitisB}(d)$. *Secondly*, modal operators are only applied to assertional knowledge (the ABox in OWL)—neither TBox nor RBox axioms are being affected by modals in our approach, as they are supposed to express universal truth.

3 CONFIDENCE AND CONFIDENCE INTERVALS

We address the *confidence* of an asserted (medical) statement (Schulz et al., 2014) through *graded* modalities applied to propositional formulae: E (*excluded*), U (*unlikely*), N (*not excluded*), L (*likely*), and C (*confirmed*). For various (technical) reasons, we add a *wildcard* modality ? (*unknown*), a complementary *failure* modality ! (*error*), plus two further modalities to syntactically state definite truth and falsity: $\top$ (*true*) and $\bot$ (*false*).[1]

Let $\triangle$ now denotes the set of all modalities:

$$\triangle := \{?, !, \top, \bot, E, U, N, L, C\}$$

A *measure function*

$$\mu : \triangle \mapsto [0,1] \times [0,1]$$

[1]We also call $\top$ and $\bot$ *propositional* modals as they lift propositional statements to the modal domain. We refer to ? and ! as *completion* modals since they complete the modal hierarchy by adding unique most general and most specific elements (see Section 4.3).

is a mapping which returns the associated *confidence interval* $\mu(\delta) = [l,h]$ for a modality from $\delta \in \triangle$ ($l \leq h$). We presuppose that

- $\mu(?) = [0,1]$
- $\mu(\top) = [1,1]$
- $\mu(\bot) = [0,0]$
- $\mu(!) = \emptyset$[2]

In addition, we define two disjoint subsets of $\triangle$, called

$$\underline{1} := \{\top, C, L, N\}$$

and

$$\underline{0} := \{\bot, E, U\}$$

and again make a presupposition: the confidence intervals for modals from $\underline{1}$ *end* in 1, whereas the confidence intervals for $\underline{0}$ modals always *start* with 0. It is worth noting that we do *not* make use of μ in the syntax of the modal language (for which we employ the modalities from $\triangle$), but in the semantics when dealing with the satisfaction relation of the model theory (see Section 4).

We have talked about *confidence intervals* now several times without saying what we actually mean by this. Suppose that a physician says that it is *confirmed* (= C) that patient p suffers from disease d, for a set of observed symptoms (or evidence) $S = \{S_1, \ldots, S_k\}$: $C\mathit{suffersFrom}(p,d)$.

Assuming that a different patient p' shows the same symptoms S (and only S, and perhaps further symptoms which are, however, *independent* from S), we would assume that the same doctor would diagnose $C\mathit{suffersFrom}(p',d)$.

Even an other, but similar trained physician is supposed to grade the two patients *similarly*. This similarity which originates from patients showing the same symptoms and from physicians being taught at the same medical school is addressed by confidence *intervals* and not through a *single* (posterior) probability, as there are still variations in diagnostic capacity and daily mental state of the physician. By using intervals (instead of single values), we can usually reach a consensus among people upon the *meaning* of gradation words, even though the low/high values of the confidence interval for, e.g., *confirmed* might depend on the context.

Being a bit more theoretic, we define a *confidence interval* as follows. Assume a *Bernoulli experiment* (Krengel, 2003) that involves a large set of n patients P, sharing the same symptoms S. W.r.t. our example, we would like to know whether $\mathit{suffersFrom}(p,d)$ or $\neg\mathit{suffersFrom}(p,d)$ is the case for every patient $p \in P$, sharing S. Given a Bernoulli trials sequence $\vec{X} = (X_1, \ldots, X_n)$ with indicator random variables $X_i \in \{0,1\}$ for a patient sequence $(p_1, \ldots, p_n)$, we can approximate the *expected value* E for *suffersFrom* being *true*, given disease d and background symptoms S by the *arithmetic mean* A:

$$\mathrm{E}[\vec{X}] \approx \mathrm{A}[\vec{X}] = \frac{\sum_{i=1}^{n} X_i}{n}$$

Due to the *law of large numbers*, we expect that if the number of elements in a trials sequence goes to infinity, the arithmetic mean will coincide with the expected value:

$$\mathrm{E}[\vec{X}] = \lim_{n \to \infty} \frac{\sum_{i=1}^{n} X_i}{n}$$

Clearly, the arithmetic mean for each new *finite* trials sequence is different, but we can try to *locate* the expected value within an interval around the arithmetic mean:

$$\mathrm{E}[\vec{X}] \in [\mathrm{A}[\vec{X}] - \varepsilon_1, \mathrm{A}[\vec{X}] + \varepsilon_2]$$

For the moment, we assume $\varepsilon_1 = \varepsilon_2$, so that $\mathrm{A}[\vec{X}]$ is in the center of this interval which we will call from now on *confidence interval*.

Coming back to our example and assuming $\mu(C) = [0.9, 1]$, $C\mathit{suffersFrom}(p,d)$ can be read as being true in 95% of all cases *known* to the physician, involving patients p potentially having disease d and sharing the same prior symptoms (evidence) $S_1, \ldots, S_k$:

$$\frac{\sum_{p \in P} \mathrm{Prob}(\mathit{suffersFrom}(p,d) | S)}{n} \approx 0.95$$

The variance of $\pm 5\%$ is related to varying diagnostic capabilities between (comparative) physicians, daily mental form, undiscovered important symptoms or examinations which have not been carried out (e.g., lab values), or perhaps even by the physical stature of the patient (crooked vs. upright) which unconsciously affects the final diagnosis, etc, as elaborated above. Thus the individual modals from $\triangle$ express (via μ) different forms of the physician's *confidence*, depending on the set of already acquired symptoms as (potential) explanations for a specific disease.

4 MODEL THEORY AND NEGATION NORMAL FORM

Let $\mathcal{C}$ denote the set of constants that serve as the arguments of a relation instance. For instance, in

[2]Recall that intervals are (usually infinite) sets of real numbers, together with an ordering relations (e.g., $<$ or $\leq$) over the elements, thus $\emptyset$ is a perfect, although degraded interval.

an RDF/OWL setting, C would exclusively consist of XSD atoms, blank nodes, and URIs/IRIs. In order to define basic n-ary propositional formulae (ground atoms), let $p(\vec{c})$ abbreviates $p(c_1,\ldots,c_n)$, for $c_1,\ldots,c_n \in C$, given $length(\vec{c}) = n$. In case the number of arguments does not matter, we sometimes simply write p, instead of, e.g., $p(c,d)$ or $p(\vec{c})$. As before, we assume $\triangle = \{?,!,\top,\bot,E,U,N,L,C\}$. We inductively define the set of *well-formed formulae* ϕ of our modal language as follows:

$$\phi ::= p(\vec{c}) \mid \neg\phi \mid \phi \wedge \phi' \mid \phi \vee \phi' \mid \triangle\phi$$

4.1 Simplification and Normal Form

We now syntactically *simplify* the set of well-formed formulae ϕ by restricting the uses of *negation* and *modalities* to the level of propositional letters π:

$$\pi ::= p(\vec{c}) \mid \neg p(\vec{c})$$
$$\phi ::= \pi \mid \triangle\pi \mid \phi \wedge \phi' \mid \phi \vee \phi'$$

The design of this language is driven by two main reasons: *firstly*, we want to effectively implement the logic (in our case, in *HFC*), and *secondly*, the application of the below semantic-preserving simplification rules in an offline pre-processing step makes the implementation easier and guarantees a more efficient runtime system. To address negation, we first need the notion of a *complement* modal δ^{C} for every $\delta \in \triangle$, where

$$\mu(\delta^{\mathsf{C}}) := \mu(\delta)^{\mathsf{C}} = \mu(?) \setminus \mu(\delta) = [0,1] \setminus \mu(\delta)$$

I.e., $\mu(\delta^{\mathsf{C}})$ is defined as the complementary interval of $\mu(\delta)$ (within the bounds of $[0,1]$, of course). For example, E and N (*excluded, not excluded*) or ? and ! (*unknown, error*) are already existing complementary modals.

We also require *mirror* modals δ^{M} for every $\delta \in \triangle$ whose confidence interval $\mu(\delta^{\mathsf{M}})$ is derived by "mirroring" $\mu(\delta)$ to the opposite side of the confidence interval, either to the left or to the right:

$$\textbf{if}\ \mu(\delta) = [l,1]\ \textbf{then}\ \mu(\delta^{\mathsf{M}}) := [0,1-l]$$
$$\textbf{if}\ \mu(\delta) = [0,h]\ \textbf{then}\ \mu(\delta^{\mathsf{M}}) := [1-h,1]$$

It is easy to see that these two equations can be unified and *generalized*[3]:

$$\textbf{if}\ \mu(\delta) = [l,h]\ \textbf{then}\ \mu(\delta^{\mathsf{M}}) := [1-h,1-l]$$

For example, E and C (*excluded, confirmed*) or $\top$ and $\bot$ (*top, bottom*) are mirror modals. In order to transform ϕ into its *negation normal form*, we need to apply simplification rules a finite number of times (until rules are no longer applicable). We depict those rules by using the $\vdash$ relation, read as *formula* $\vdash$ *simplified formula* (ε = empty word):

1. $?\phi \vdash \varepsilon$ *% $?\phi$ is not informative at all*
2. $\neg\neg\phi \vdash \phi$
3. $\neg(\phi \wedge \phi') \vdash \neg\phi \vee \neg\phi'$
4. $\neg(\phi \vee \phi') \vdash \neg\phi \wedge \neg\phi'$
5. $\neg\triangle\phi \vdash \triangle^{\mathsf{C}}\phi$ (example: $\neg E\phi = E^{\mathsf{C}}\phi = N\phi$)
6. $\triangle\neg\phi \vdash \triangle^{\mathsf{M}}\phi$ (example: $E\neg\phi = E^{\mathsf{M}}\phi = C\phi$)

Clearly, the mirror modals δ^{M} ($\delta \in \triangle$) are not necessary as long as we explicitly allow for negated statements (which we do), and thus case 6 can, in principle, be dropped.

What is the result of simplifying $\triangle(\phi \wedge \phi')$ and $\triangle(\phi \vee \phi')$? Let us start with the former case and consider as an example the statement about an engine that *a mechanical failure m <u>and</u> an electrical failure e is <u>confirmed</u>*: $C(m \wedge e)$. It seems *plausible* to simplify this expression to $Cm \wedge Ce$. Commonsense tells us furthermore that neither Em nor Ee is compatible with this description (we should be alarmed if, e.g., both Cm and Em happen to be the case).

Now consider the "opposite" statement $E(m \wedge e)$ which must *not* be rewritten to $Em \wedge Ee$, as *either Cm or Ce* is well *compatible* with $E(m \wedge e)$. Instead, we rewrite this kind of "negated" statement as $Em \vee Ee$, and this works fine with either Cm or Ce.

In order to address the other modal operators, we generalize these *plausible* inferences by making a distinction between $\underline{0}$ and $\underline{1}$ modals (cf. Section 3):

7a. $\underline{0}(\phi \wedge \phi') \vdash \underline{0}\phi \vee \underline{0}\phi'$

7b. $\underline{1}(\phi \wedge \phi') \vdash \underline{1}\phi \wedge \underline{1}\phi'$

Let us now focus on disjunction inside the scope of a modal operator. As we do allow for the full set of Boolean operators, we are allowed to deduce

8. $\triangle(\phi \vee \phi') \vdash \triangle(\neg(\neg(\phi \vee \phi'))) \vdash \triangle(\neg(\neg\phi \wedge \neg\phi')) \vdash \triangle^{\mathsf{M}}(\neg\phi \wedge \neg\phi')$

This is, again, a conjunction, so we apply schemas 7a and 7b, giving us

8a. $\underline{0}(\phi \vee \phi') \vdash \underline{0}^{\mathsf{M}}(\neg\phi \wedge \neg\phi') \vdash \underline{1}(\neg\phi \wedge \neg\phi') \vdash \underline{1}\neg\phi \wedge \underline{1}\neg\phi' \vdash \underline{1}^{\mathsf{M}}\phi \wedge \underline{1}^{\mathsf{M}}\phi' \vdash \underline{0}\phi \wedge \underline{0}\phi'$

8b. $\underline{1}(\phi \vee \phi') \vdash \underline{1}^{\mathsf{M}}(\neg\phi \wedge \neg\phi') \vdash \underline{0}(\neg\phi \wedge \neg\phi') \vdash \underline{0}\neg\phi \vee \underline{0}\neg\phi' \vdash \underline{0}^{\mathsf{M}}\phi \vee \underline{0}^{\mathsf{M}}\phi' \vdash \underline{1}\phi \vee \underline{1}\phi'$

Note how the modals from $\underline{0}$ in 7a and 8a act as a kind of *negation* operator to turn the logical operators into their counterparts, similar to de *Morgan's law*.

The final case considers two consecutive modals:

[3] This construction procedure comes in handy when dealing with *in-the-middle* modals, such as *fifty-fifty* or *perhaps*, whose confidence intervals neither touch 0 nor 1. Such modals have a *real* background in (medical) diagnosis.

9. $\delta' \delta'' \phi \vdash (\delta' \circ \delta'')\phi$

We interpret the $\circ$ operator as a kind of *function composition*, leading to a new modal δ which is the result of $\delta' \circ \delta''$. We take a liberal stance here of what the result is, but indicate that it depends on the domain and, again, plausible inferences we like to capture. The $\circ$ operator will probably be different from the related operation $\odot$ which is used in Section 5.3.4.

4.2 Model Theory

In the following, we extend the standard definition of modal (Kripke) frames and models (Blackburn et al., 2001) for the *graded* modal operators from $\triangle$ by employing the confidence function μ and focussing on the minimal definition for ϕ. A *frame* $\mathcal{F}$ for the probabilistic modal language is a pair

$$\mathcal{F} = \langle \mathcal{W}, \triangle \rangle$$

where $\mathcal{W}$ is a non-empty set of *worlds* (or *situations, states, points, vertices, etc.*) and $\triangle$ a family of binary relations over $\mathcal{W} \times \mathcal{W}$, called *accessibility relations*. In the following, we *overload* $\delta \in \triangle$ below in that we let δ both refer to the modal in the syntax as well as to the accessibility relation R_δ in the semantics.

A *model* $\mathcal{M}$ for the probabilistic modal language is a triple

$$\mathcal{M} = \langle \mathcal{F}, \mathcal{V}, \mu \rangle$$

such that $\mathcal{F}$ is a *frame*, $\mathcal{V}$ is a *valuation*, assigning each proposition ϕ a subset of $\mathcal{W}$, viz., the set of worlds in which ϕ holds, and μ is a mapping, returning the confidence interval for a given modality from $\triangle$. Note that we only require a definition for μ in $\mathcal{M}$ (the model, but *not* in the frame), as $\mathcal{F}$ represents the relational structure without interpreting the edge labeling (the modal names) of the graph.

The *satisfaction relation* $\models$, given a model $\mathcal{M}$ and a specific world w is inductively defined over the set of well-formed formulae in *negation normal form* (remember $\pi ::= p(\vec{c}) \mid \neg p(\vec{c})$):

1. $\mathcal{M}, w \models p(\vec{c})$ **iff** $w \in \mathcal{V}(p(\vec{c}))$ **and** $w \notin \mathcal{V}(\neg p(\vec{c}))$
2. $\mathcal{M}, w \models \neg p(\vec{c})$ **iff** $w \in \mathcal{V}(\neg p(\vec{c}))$ **and** $w \notin \mathcal{V}(p(\vec{c}))$
3. $\mathcal{M}, w \models \phi \wedge \phi'$ **iff** $\mathcal{M}, w \models \phi$ **and** $\mathcal{M}, w \models \phi'$
4. $\mathcal{M}, w \models \phi \vee \phi'$ **iff** $\mathcal{M}, w \models \phi$ **or** $\mathcal{M}, w \models \phi'$
5. **for all** $\delta \in \triangle$: $\mathcal{M}, w \models \delta\pi$ **iff**

$$\frac{\#\{u \mid (w,u) \in \delta \textbf{ and } \mathcal{M}, u \models \pi\}}{\# \cup_{\delta' \in \triangle} \{u \mid (w,u) \in \delta'\}} \in \mu(\delta)$$

The last case of the satisfaction relation addresses the modals: for a world w, we look for the successor states u that are directly reachable via δ and in which π holds, and divide the number of such states ($\# \cdot$) by the number of all worlds that are reachable from w in the denominator. This number, lying between 0 and 1, is then required to be an element of the confidence interval $\mu(\delta)$ of δ in order to satisfy $\delta\pi$, given $\mathcal{M}, w$.

It is worth noting that the satisfaction relation above differs from the standard definition in its handling of $\mathcal{M}, w \models \neg p(\vec{c})$, as negation is *not* interpreted through the *absence* of $p(\vec{c})$ ($\mathcal{M}, w \not\models p(\vec{c})$), but through the *existence* of $\neg p(\vec{c})$. This treatment addresses the *open-world* nature in OWL and the evolvement of a (medical) domain over time.

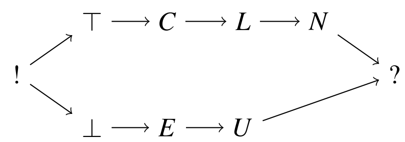
We also note that the definition of the satisfaction relation for modalities (last clause) is related to the *possibility operators* $M_k \cdot$ (= $\Diamond^{\geq k} \cdot$; $k \in \mathbb{N}$) introduced by (Fine, 1972) and *counting modalities* $\cdot \geq n$ (Areces et al., 2010), used in modal logic characterizations of *description logics* with *cardinality* restrictions.

4.3 Well-Behaved Frames

As we will see later, it is handy to assume that the graded modals are arranged in a kind of hierarchy—the more we move along the arrows in the hierarchy, the more a statement ϕ in the scope of a modal $\delta \in \triangle$ becomes *uncertain*. In order to address this, we slightly extend the notion of a *frame* by a third component $\preceq \subseteq \triangle \times \triangle$, a partial order (i.e., a reflexive, antisymmetric, and transitive binary relation) between modalities:

$$\mathcal{F} = \langle \mathcal{W}, \triangle, \preceq \rangle$$

Let us consider the following modal hierarchy that we build from the set $\triangle$ of already introduced modals (cf. Figure 1):



This graphical representation is just a compact way to specify a set of 33 binary relation instances over $\triangle \times \triangle$, such as $\top \preceq \top$, $\top \preceq N$, $C \preceq N$, $\perp \preceq ?$, or $! \preceq ?$. The above mentioned form of uncertainty is expressed by the measure function μ in that the associated confidence intervals become larger:

$$\textbf{if } \delta \preceq \delta' \textbf{ then } \mu(\delta) \subseteq \mu(\delta')$$

In order to arrive at a proper and intuitive model-theoretic semantics which mirrors intuitions such as **if** ϕ *is confirmed* ($C\phi$) **then** ϕ *is likely* ($L\phi$), we will focus here on *well-behaved* frames $\mathcal{F}$ which enforce the existence of edges in $\mathcal{W}$, given $\preceq$ and $\delta, \delta^{\uparrow} \in \triangle$:

$$\textbf{if } (w,u) \in \delta \textbf{ and } \delta \preceq \delta^{\uparrow} \textbf{ then } (w,u) \in \delta^{\uparrow}$$

However, by imposing this constraint, we also need to adapt the last case of the satisfiability relation from Section 4.2 above:

5. **for all** $\delta \in \triangle$: $\mathcal{M}, w \models \delta\pi$ **iff**

$$\frac{\# \cup_{\delta^{\uparrow} \succeq \delta} \{u \mid (w,u) \in \delta^{\uparrow} \textbf{ and } \mathcal{M}, u \models \pi\}}{\# \cup_{\delta' \in \triangle} \{u \mid (w,u) \in \delta'\}} \in \mu(\delta)$$

Not only are we scanning for edges (w,u) labeled with δ and for successor states u of w in which π holds in the numerator (original definition), but also take into account edges marked with more general modals $\delta^{\uparrow}$: $\delta^{\uparrow} \succeq \delta$. This mechanism implements a kind of *built-in model completion* that is not necessary in ordinary modal logics as they deal with only a *single* relation (viz., unlabeled arcs).

5 ENTAILMENT RULES

We now turn our attention, again, to the syntax of our language and to the syntactic consequence relation. This section addresses a restricted subset of entailment rules which will unveil new (or implicit) knowledge from already existing graded statements. Recall that these kind of statements (in negation normal form) are a consequence of the application of simplification rules as depicted in Section 4.1. Thus, we assume a *pre-processing step* here that "massages" more complex statements that arise from a representation of graded (medical) statements in *natural language*. The entailments which we will present in a moment can either be *directly* implemented in a *tuple*-based reasoner, such as *HFC* (Krieger, 2013), or in *triple*-based engines (e.g., *Jena* (Carroll et al., 2004) or *OWLIM* (Bishop et al., 2011)) which need to *reify* the medical statements in order to be compliant with the RDF triple model.

5.1 Modal Entailments

The entailments presented in this section deal with *plausible* inference centered around modals $\delta, \delta' \in \triangle$ which are, in part, also addressed in (Schulz et al., 2014) in a pure OWL setting. We use the implication sign $\rightarrow$ to depict the entailment rules

$$lhs \rightarrow rhs$$

which act as *completion* (or *materialization*) rules the way as described in, e.g., (Hayes, 2004) and (ter Horst, 2005), and used in today's *semantic repositories* (e.g., *OWLIM*). We sometimes even use the biconditional $\leftrightarrow$ to address that the LHS and the RHS are semantically equivalent, but will indicate the direction that should be used in a practical setting. As before, we define

$$\pi ::= p(\vec{c}) \mid \neg p(\vec{c})$$

We furthermore assume that for every modal $\delta \in \triangle$, a *complement* modal δ^{C} and a *mirror* modal δ^{M} exist (cf. Section 4.1).

5.1.1 Lift

$$(\mathsf{L}) \quad \pi \leftrightarrow \top\pi$$

This rule interprets propositional statements as special modal formulae. It might be dropped and can be seen as a pre-processing step. We have used it in the Hepatitis example above. Usage: left-to-right direction.

5.1.2 Generalize

$$(\mathsf{G}) \quad \delta\pi \wedge \delta \preceq \delta' \rightarrow \delta'\pi$$

This rule schema can be instantiated in various ways, using the modal hierarchy from Section 4.3, e.g., $\top\pi \rightarrow C\pi$, $C\pi \rightarrow L\pi$, or $E\pi \rightarrow U\pi$. It has been used in the Hepatitis example.

5.1.3 Complement

$$(\mathsf{C}) \quad \neg\delta\pi \leftrightarrow \delta^{\mathsf{C}}\pi$$

In principle, (C) is not needed in case the statement is already in negation normal form. This schema might be useful for natural language paraphrasing (explanation). Given $\triangle$, there are four possible instantiations: $E\pi \leftrightarrow \neg N\pi$, $N\pi \leftrightarrow \neg E\pi$, $?\pi \leftrightarrow \neg !\pi$, and $!\pi \leftrightarrow \neg ?\pi$.

5.1.4 Mirror

$$(\mathsf{M}) \quad \delta\neg\pi \leftrightarrow \delta^{\mathsf{M}}\pi$$

Again, (D) is in principle not needed as long as the modal proposition is in negation normal form, since we do allow for negated propositional statements $\neg p(\vec{c})$. This schema might be useful for natural language paraphrasing (explanation). For $\triangle$, there are six possible instantiations, viz., $E\pi \leftrightarrow C\neg\pi$, $C\pi \leftrightarrow E\neg\pi$, $L\pi \leftrightarrow U\neg\pi$, $U\pi \leftrightarrow L\neg\pi$, $\top\pi \leftrightarrow \bot\neg\pi$, and $\bot\pi \leftrightarrow \top\neg\pi$.

5.1.5 Uncertainty

$$\text{(U)} \quad \delta\pi \wedge \neg\delta\pi \leftrightarrow \delta\pi \wedge \delta^{\mathsf{C}}\pi \leftrightarrow ?\pi$$

The *co-occurrence* of $\delta\pi$ and $\neg\delta\pi$ does *not* imply logical *inconsistency* (propositional case: $\pi \wedge \neg\pi$), but leads to complete *uncertainty* about the validity of π. Remember that $\mu(?) = \mu(\delta) \uplus \mu(\delta^{\mathsf{C}}) = [0,1]$:

```
       0                 1
 μ : |—δC—|———δ———|
         π          π
```

Usage: left-to-right direction.

5.1.6 Negation

$$\text{(N)} \quad \delta(\pi \wedge \neg\pi) \leftrightarrow \delta\pi \wedge \delta\neg\pi \leftrightarrow \delta\pi \wedge \delta^{\mathsf{M}}\pi \leftrightarrow \delta^{\mathsf{M}}\neg\pi \wedge \delta^{\mathsf{M}}\pi \leftrightarrow \delta^{\mathsf{M}}(\pi \wedge \neg\pi)$$

(N) shows that $\delta(\pi \wedge \neg\pi)$ can be formulated equivalently by using the mirror modal δ^{M}:

```
       0                     1
 μ : |—δM—|———|—δ—|
       π∧¬π        π∧¬π
```

In general, (N) is *not* the modal counterpart of the *law of non-contradiction*, as $\pi \wedge \neg\pi$ is usually afflicted by uncertainty, meaning that from $\delta(\pi \wedge \neg\pi)$, we can *not* infer that $\pi \wedge \neg\pi$ is the case for the concrete example in question (recall the intention behind the confidence intervals; cf. Section 3). There is one notable exception, involving the $\top$ and $\bot$ modals. This is formulated by the next entailment rule.

5.1.7 Error

$$\text{(E)} \quad \top(\pi \wedge \neg\pi) \leftrightarrow \bot(\pi \wedge \neg\pi) \rightarrow !(\pi \wedge \neg\pi) \leftrightarrow !\pi$$

(E) *is* the modal counterpart of the *law of non-contradiction* (note: $\bot^{\mathsf{M}} = \top, \top^{\mathsf{M}} = \bot, !^{\mathsf{M}} = !$). For this reason and *by definition*, the *error* (or *failure*) modal ! from Section 3 comes into play here. The modal ! can serve as a hint to either stop a computation the first time it occurs, or to continue reasoning and to syntactically memorize the ground literal π. Usage: left-to-right direction.

5.2 Subsumption Entailments

As before, we define two subsets of $\triangle$, called $\underline{1} = \{\top, C, L, N\}$ and $\underline{0} = \{\bot, E, U\}$, thus $\underline{1}$ and $\underline{0}$ effectively become

$$\underline{1} = \{\top, C, L, N, U^{\mathsf{C}}\} \qquad \underline{0} = \{\bot, U, E, C^{\mathsf{C}}, L^{\mathsf{C}}, N^{\mathsf{M}}\}$$

due to the use of complement modals δ^{C} and mirror modals δ^{M} for every base modal $\delta \in \triangle$ and by assuming that $E = N^{\mathsf{C}}$, $E = C^{\mathsf{M}}$, $U = L^{\mathsf{M}}$, and $\bot = \top^{\mathsf{M}}$, together with the four "opposite" cases.

Now, let $\sqsubseteq$ abbreviate relation subsumption as known from description logics and realized through `rdfs:subClassOf` and `rdfs:subPropertyOf`. Given this, we define two further very practical and plausible modal entailments which can be seen as the modal extension of the entailment rules (rdfs9) and (rdfs7) for classes and properties in RDFS (Hayes, 2004):

$$\text{(S1)} \quad \underline{1}p(\vec{c}) \wedge p \sqsubseteq q \rightarrow \underline{1}q(\vec{c})$$

$$\text{(S0)} \quad \underline{0}q(\vec{c}) \wedge p \sqsubseteq q \rightarrow \underline{0}p(\vec{c})$$

Note how the use of p and q switches in the antecedent and the consequent, even though $p \sqsubseteq q$ holds in both cases. Note further that propositional statements π are restricted to the positive case $p(\vec{c})$ and $q(\vec{c})$, as their negation in the antecedent will not lead to any valid entailments.

Here are four *instantiations* of (S0) and (S1) for the unary and binary case (remember, $C \in \underline{1}$ and $E \in \underline{0}$):

$\mathsf{ViralHepatitisB} \sqsubseteq \mathsf{ViralHepatitis} \wedge C\mathsf{ViralHepatitisB}(x) \rightarrow C\mathsf{ViralHepatitis}(x)$

$\mathsf{ViralHepatitis} \sqsubseteq \mathsf{Hepatitis} \wedge E\mathsf{Hepatitis}(x) \rightarrow E\mathsf{ViralHepatitis}(x)$

$\mathsf{deeplyEnclosedIn} \sqsubseteq \mathsf{containedIn} \wedge C\mathsf{deeplyEnclosedIn}(x,y) \rightarrow C\mathsf{containedIn}(x,y)$

$\mathsf{superficiallyLocatedIn} \sqsubseteq \mathsf{containedIn} \wedge E\mathsf{containedIn}(x,y) \rightarrow E\mathsf{superficiallyLocatedIn}(x,y)$

5.3 Extended RDFS & OWL Entailments

In this section, we will consider further entailment rules for RDFS (Hayes, 2004) and a restricted subset of OWL (ter Horst, 2005; Motik et al., 2012). Remember that modals only head positive and negative propositional letters π, not TBox or RBox axioms. Concerning the original entailment rules, we will distinguish *four principal cases* to which the extended rules belong (we will only consider the unary and binary case here as used in description logics/OWL):

1. TBox and RBox axiom schemas will not undergo a modal extension;
2. rules get extended in the antecedent;
3. rules take over modals from the antecedent to the consequent;
4. rules aggregate several modals from the antecedent in the consequent.

We will illustrate the individual cases in the following subsections with examples by using a kind of description logic rule syntax. Clearly, the set of extended entailments depicted here is *not complete*.

5.3.1 Case-1: No Modals

Entailment rule (rdfs11) from (Hayes, 2004) deals with class subsumption: $\mathsf{C} \sqsubseteq \mathsf{D} \wedge \mathsf{D} \sqsubseteq \mathsf{E} \rightarrow \mathsf{C} \sqsubseteq \mathsf{E}$. As this is a terminological axiom schema, the rule stays *constant* in the modal domain. Example rule instantiation:

$\mathsf{ViralHepatitisB} \sqsubseteq \mathsf{ViralHepatitis} \wedge$
$\mathsf{ViralHepatitis} \sqsubseteq \mathsf{Hepatitis} \rightarrow$
$\mathsf{ViralHepatitisB} \sqsubseteq \mathsf{Hepatitis}$

5.3.2 Case-2: Modals on LHS, No Modals on RHS

The following original rule (rdfs3) from (Hayes, 2004) imposes a range restriction on objects of binary ABox relation instances: $\forall \mathsf{P.C} \wedge \mathsf{P(x,y)} \rightarrow \mathsf{C(y)}$. The extended version needs to address the ABox proposition in the antecedent (*don't care* modal δ), but must not change the consequent (even though we always use the $\top$ modality here—the range restriction C(y) is always true, independent of the uncertainty of P(x, y); cf. Section 2 example):

$$\text{(Mrdfs3)} \quad \forall \mathsf{P.C} \wedge \delta \mathsf{P(x,y)} \rightarrow \top \mathsf{C(y)}$$

Example rule instantiation:

$\forall \mathsf{suffersFrom.Disease} \wedge L\mathsf{suffersFrom}(x,y) \rightarrow$
$\top \mathsf{Disease}(y)$

5.3.3 Case-3: Keeping LHS Modals on RHS

Inverse properties switch their arguments (ter Horst, 2005) as described by (rdfp8): $\mathsf{P} \equiv \mathsf{Q}^- \wedge \mathsf{P(x,y)} \rightarrow \mathsf{Q(y,x)}$. The extended version simply keeps the modal operator:

$$\text{(Mrdfp8)} \quad \mathsf{P} \equiv \mathsf{Q}^- \wedge \delta \mathsf{P(x,y)} \rightarrow \delta \mathsf{Q(y,x)}$$

Example rule instantiation:

$\mathsf{containedIn} \equiv \mathsf{contains}^- \wedge C\mathsf{containedIn}(x,y) \rightarrow$
$C\mathsf{contains}(y,x)$

5.3.4 Case-4: Aggregating LHS Modals on RHS

Now comes the most interesting case of modalized RDFS & OWL entailment rules, that offers several possibilities on a varying scale between *skeptical* and *credulous* entailments, depending on the degree of uncertainty, as expressed by the measuring function μ of the modal operator. Consider the original rule (rdfp4) from (ter Horst, 2005) for transitive properties:

$\mathsf{P}^+ \sqsubseteq \mathsf{P} \wedge \mathsf{P(x,y)} \wedge \mathsf{P(y,z)} \rightarrow \mathsf{P(x,z)}$.

Now, how does the modal on the RHS of the extended rule look like, depending on the two LHS modals? There are several possibilities. By operating directly on the *modal hierarchy*, we are allowed to talk about, e.g., the *least upper bound* or the *greatest lower bound* of δ' and δ''. When taking the associated *confidence intervals* into account, we might play with the low and high numbers of the intervals, say, by applying min/max, the *arithmetic mean* or even by *multiplying* the corresponding numbers.

Let us first consider the general rule from which more specialized versions can be derived, simply by instantiating the combination operator $\odot$:

$$\text{(Mrdfp4)} \quad \mathsf{P}^+ \sqsubseteq \mathsf{P} \wedge \delta' \mathsf{P(x,y)} \wedge \delta'' \mathsf{P(y,z)} \rightarrow (\delta' \odot \delta'')\mathsf{P(x,z)}$$

Here is an instantiation of Mrdfp4, dealing with the transitive relation contains from above, assuming that $\odot$ reduces to the *least upper bound* (i.e., $C \odot L = L$):

$C\mathsf{contains}(x,y) \wedge L\mathsf{contains}(y,z) \rightarrow$
$L\mathsf{contains}(x,z)$

What is the general result of $\delta' \odot \delta''$? It depends, probably both on the application domain and the *epistemic commitment* one is willing to accept about the "meaning" of gradation words/modal operators. To enforce that $\odot$ is at least both *commutative* and *associative* (as is the least upper bound) is probably a good idea, making the sequence of modal clauses *order independent*. And to work on the modal hierarchy instead of combining low/high numbers of the corresponding intervals is probably a good decision for forward chaining engines, as the latter strategy might introduce *new* individuals through operations such as multiplication, thus posing a problem for the implementation of the generalization schema (G) (see Section 5.1.2).

5.4 Custom Entailments: An Example from the Medical Domain

Consider that Hepatitis B is an infectious disease

$\mathsf{ViralHepatitisB} \sqsubseteq \mathsf{InfectiousDisease} \sqsubseteq \mathsf{Disease}$

and note that there exist vaccines against it. Assume that the liver l of patient p quite hurts

$C\mathsf{hasPain(p, l)}$,

but p has been definitely vaccinated against Hepatitis B before:

$\top\mathsf{vaccinatedAgainst(p, ViralHepatitisB)}$.

We apply OWL2-like punning here when using the class ViralHepatitisB (*not* an instance), as the second argument of vaccinatedAgainst; cf. (Golbreich and Wallace, 2012).

Given that p received a vaccination, the following custom rule will *not* fire (x, y below are now universally-quantified variables; z an existentially-quantified RHS-only variable):

$$\top \mathsf{Patient}(x) \wedge \top \mathsf{Liver}(y) \wedge C \mathsf{hasPain}(x,y) \wedge U \mathsf{vaccinatedAgainst}(x, \mathsf{ViralHepatitisB}) \rightarrow N \mathsf{ViralHepatitisB}(z) \wedge N \mathsf{suffersFrom}(x,z)$$

Now assume another person p' that is pretty sure (s)he was never vaccinated:

$$E \mathsf{vaccinatedAgainst}(\mathsf{p}', \mathsf{ViralHepatitisB})$$

Given the above custom rule, we are allowed to infer that (h instantiation of z)

$$N \mathsf{ViralHepatitisB}(\mathsf{h}) \wedge N \mathsf{suffersFrom}(\mathsf{p}', \mathsf{h})$$

The subclass axiom from above thus assigns

$$N \mathsf{InfectiousDisease}(\mathsf{h})$$

so that we can query for patients for whom an infectious disease is *not excluded* (= N), in order to initiate appropriate methods (e.g., further medical investigations).

5.5 Implementing Modal Entailments

The negation normal form from Section 4.1 makes it relatively easy to implement entailment rules involving modalized propositional letters of the form $\delta \pm p(\vec{c})$. $\pm$ is a *polarity value* as known from situation theory (Devlin, 2006) in order to make negative property assertions available in the object language.

We have implemented a modalized extension of the RDFS and OWL rule sets (Hayes, 2004; ter Horst, 2005) by employing the tuple-based rule engine *HFC* (Krieger, 2012; Krieger, 2013). Without loss of generality, let us focus here on the positive case for the three binary entailment schemas from Section 5.3.2, 5.3.3, and 5.3.4 and their *HFC* rule representation, as negation inside the scope of a modal can be rewritten using the mirror modal, thus turning the quintuple into a quad (rule variables start with a ?):

(Mrdfs3) $\forall \mathsf{P.C} \wedge \delta \mathsf{P(x,y)} \rightarrow \top \mathsf{C(y)}$

```
?p rdfs:range ?c
?modal ?x ?p ?y
->
mod:T ?y rdf:type ?c
```

(Mrdfp8) $\mathsf{P} \equiv \mathsf{Q}^- \wedge \delta \mathsf{P(x,y)} \rightarrow \delta \mathsf{Q(y,x)}$

```
?p owl:inverseOf ?q
?modal ?x ?p ?y
->
?modal ?y ?q ?x
```

(Mrdfp4) $\mathsf{P}^+ \sqsubseteq \mathsf{P} \wedge \delta \mathsf{P(x,y)} \wedge \delta' \mathsf{P(y,z)} \rightarrow (\delta \odot \delta') \mathsf{P(x,z)}$

```
?p rdf:type owl:TransitiveProperty
?modal1 ?x ?p ?y
?modal2 ?y ?p ?z
->
?modal ?x ?p ?z
@action
?modal = CombineModals ?modal1 ?modal2
```

Triple-based engines, such as *OWLIM* clearly need to reify such extended descriptions (expensive; no termination guarantee). Even more important, additional tests going beyond simple symbol matching and function calls, such as `CombineModals` (the equivalent to $\odot$ in the abstract syntax) in the *HFC* version of (Mrdfp4) above, are rarely available in today's RDFS/OWL reasoning engines, thus making it impossible for them to implement such modal entailments.

We finally describe how the implementation of the generalization schema (G) (Section 5.1.2) works. As explained in Section 4.3, the modal operators δ are arranged in a modal hierarchy that is based on the inclusion of their confidence intervals $\mu(\delta)$. This hierarchy is realized in OWL through a subclass hierarchy, using `rdfs:subClassOf` to implement $\preceq$:

(G) $\delta \mathsf{P(x,y)} \wedge \delta \preceq \delta' \rightarrow \delta' \mathsf{P(x,y)}$

```
?modal1 ?x ?p ?y
?modal1 rdfs:subClassOf ?modal2
->
?modal2 ?x ?p ?z
```

6 A FOURTH KIND OF MODALS

The two modalities $\Box$ and $\Diamond$ from standard modal logic are often called *dual* as they can be defined in terms of each other: $\Box\phi \equiv \neg\Diamond\neg\phi$ and $\Diamond\phi \equiv \neg\Box\neg\phi$, resp. At first sight, it seems that our non-standard modal logic is missing a similar property, as we originally dealt with *five* modal operators, extended by the *propositional* modals $\top$ and $\bot$, and the *completion* modals ? and !. For every such modal δ, we can furthermore think of additional *complement* modals δ^{C} and additional *mirror* modals δ^{M} whose confidence intervals $\mu(\delta^{\mathsf{C}})$ and $\mu(\delta^{\mathsf{M}})$ can be derived from $\mu(\delta)$ (cf. Section 4.1). Some of these modals coincide with original modals from $\triangle$, others do not have a direct counterpart. However, the confidence intervals for the "anonymous" modals can be trivially computed by applying the two equations from Section 4.1.

Coming back to the question of whether dual modals exist for every $\delta \in \triangle$, we need to *simplify* $\neg\delta\neg\phi$ by applying the schemas from Section 4.1. We can either start with the inner or with the outer negation, resulting in either mirror modals or complement modals. Interestingly, the resulting confidence intervals at which we reach in the end are the *same*, and this is clearly a good point and desirable, as simplification is supposed to be an *order-independent* process:

$$
\begin{array}{ccc}
 & \neg\delta\neg\phi & \\
/ & & \backslash \\
\delta^{\mathsf{C}}\neg\phi & & \neg\delta^{\mathsf{M}}\phi \\
| & & | \\
\delta^{\mathsf{C}^{\mathsf{M}}}\phi & & \delta^{\mathsf{M}^{\mathsf{C}}}\phi
\end{array}
$$

Thus, $\delta^{\mathsf{C}^{\mathsf{M}}} \equiv \delta^{\mathsf{M}^{\mathsf{C}}}$, for every $\delta \in \triangle$ which can be shown by applying the definitions for complement and mirror modals from Section 4.1. The deeper reason why this is so is related to the inherent properties of the two operations *complementation* and *mirroring*. Contrary to complement and mirror modals, dual modals δ^{D} are either supersets or subsets of $\mu(\delta)$, i.e., if δ is a $\underline{1}$- or $\underline{0}$-modal, so is δ^{D}.

7 RELATED WORK & REMARKS

It is worth noting to state that this paper is interested in the representation of and reasoning with *uncertain assertional* knowledge, and neither in dealing with *vagueness/fuzziness* found in natural language (*very small, hot*), nor in handling *defaults* and *exceptions* in *terminological* knowledge (*penguins can't fly*).

To the best of our knowledge, the modal logic presented in this paper uses for the first time modal operators for expressing the degree of (un)certainty of propositions. These modal operators are interpreted in the model theory through confidence intervals via measure function μ. From a model point of view, our modal operators are related to *counting modalities* $\Diamond^{\geq k}$ (Fine, 1972; Areces et al., 2010). However, for $\mathcal{M}, w \models \delta\pi$ to be the case, we do *not* require a *fixed* number $k \in \mathbb{N}$ of reachable successor states (*absolute* frequency), but instead *divide* the number of worlds reached through label $\delta \in \triangle$ and in which π holds by the number of all directly reachable worlds, yielding fraction $0 \leq p \leq 1$. This number then is further constrained by requiring $p \in \mu(\delta)$ (*relative* frequency), as defined in case 5 of the satisfaction relation in Section 4.2 and extended in Section 4.3.

As (Wikipedia, 2015) precisely put it: "... *what axioms and rules must be added to the propositional calculus to create a usable system of modal logic is a matter of philosophical opinion, often driven by the theorems one wishes to prove* ...". Clearly, the logic presented here is *no* exception and its design is driven by commonsense knowledge and plausible inferences, we try to capture and generalize. In a *strict* sense, it is a non-standard modal logic in that it is *not* an instance of the *normal* modal logic $\mathbf{K} = (N) + (K)$

$(N)\ p \rightarrow \Box p$
$(K)\ \Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q)$

as the *necessitation rule* (N) and the *distribution axiom* (K) does *not* hold for *every* $\delta \in \triangle$. However, we can show that restricted generalized forms of these axioms are in fact the case for our logic ($\underline{1}^{\geq 0.5}$ are $\underline{1}$-modals whose low value is ≥ 0.5 and $\underline{0}^{\leq 0.5}$ are $\underline{0}$-modals whose high value is ≤ 0.5):

$(N1)\ p \rightarrow \underline{1}p$
$(N0)\ \neg p \rightarrow \underline{0}p$
$(K1^{\geq 0.5})\ \underline{1}^{\geq 0.5}(p \rightarrow q) \rightarrow (\underline{1}^{\geq 0.5}p \rightarrow \underline{1}^{\geq 0.5}q)$
$(K0^{\leq 0.5})\ \underline{0}^{\leq 0.5}(p \rightarrow q) \rightarrow (\underline{0}^{\leq 0.5}p \rightarrow \underline{0}^{\leq 0.5}q)$

In addition, the *well-behaved frames* condition (Section 4.3) generalizes the *seriality* condition (D) on frames and a kind of *forward monotonicity*, we would like to keep for an evolving domain, is directly related to *transitivity* (4) of the accessibility relations from $\triangle$ in $\mathcal{F}$:

$(D)\ \delta p \wedge \delta \preceq \delta' \rightarrow \delta' p$
$(4)\ \delta p \rightarrow \delta\delta p$

Several approaches to *representing and reasoning with uncertainty* have been investigated in *Artificial Intelligence*; see (Halpern, 2003) for a (biased) overview. (Halpern, 1990) was probably the first attempt of a first-order logic which unifies probability distributions over classes *and* individuals. Weaker decidable propositional formalisms such as Bayesian Networks (Pearl, 1988) and related probabilistic graphical models (Koller and Friedmann, 2009) have found their way into causal (medical) reasoning (Lucas et al., 2004). Programming languages for these kind of models exist; e.g., Alchemy for *Markov Logic Networks* (Richardson and Domingos, 2006). In Markov Logic, first-order formulae are associated with a numerical value which *softens* hard first-order constraints and a violation makes a possible world not impossible, but less probable (the higher the weight, the stronger the rule). For example, the Markov Logic rule *smoking causes cancer* with weight 1.5 (Richardson and Domingos, 2006, p. 111)

$1.5 : \forall x \,.\, smokes(x) \rightarrow hasCancer(x)$

might be *approximated* in our approach through the use of modals:

$\top\mathsf{smokes(x)} \rightarrow L\mathsf{hasCancer(x)}$

Very less so has been researched in the *Description Logic* community (as it is smaller) and little or nothing of this research has find its way into implemented description logic systems. As we focus in this paper on a modalized extension of OWL, let us

$\triangle$	*meaning*	*confidence*	*belief*	*disbelief*	*uncertainty*
!	*error*	$\emptyset$	0.5	0.5	0
$\perp$	*false*	$[0,0]$	0	1	0
E	*excluded*	$[0,0.1]$	0	0.9	0.1
U	*unlikely*	$[0,0.3]$	0	0.7	0.3
PN	*perhaps not*	$[0.4,0.5]$	0.4	0.5	0.1
FF	*fifty-fifty*	$[0.45,0.55]$	0.45	0.45	0.1
P	*perhaps*	$[0.5,0.6]$	0.5	0.4	0.1
N	*not excluded*	$[0.1,1]$	0.3	0	0.7
L	*likely*	$[0.7,1]$	0.7	0	0.3
C	*confirmed*	$[0.9,1]$	0.9	0	0.1
$\top$	*true*	$[1,1]$	1	0	0
?	*unknown*	$[0,1]$	0	0	1

Figure 2: Representation of modal operators from $\triangle$ (incl. three *in-the-middle* modals) in terms of *opinions* in Subjective Logic. The confidence intervals for the five initial modals roughly coincide with the numbers depicted in Figure 1.

review here some of the work carried out in description logics. (Heinsohn, 1993) and (Jaeger, 1994) consider *uncertainty* in $\mathcal{ALC}$ concept hierarchies, plus concept typing of individuals (unary relations) in different ways (probability values vs. intervals; conditional probabilities in TBox vs. TBox+ABox). They do not address uncertain binary (or even *n*-ary) relations. (Tresp and Molitor, 1998) investigates *vagueness* in $\mathcal{ALC}$ concept descriptions to address statements, such as *the patient's temperature is high*, but also for determining membership degree ($\overline{38.5}$ °C). This is achieved through *membership manipulators* which are functions, returning a truth value between 0 and 1, thus deviating from a two-valued logic. (Straccia, 2001) defines a *fuzzy* extension of $\mathcal{ALC}$, based on Zadeh's *Fuzzy Logic*. As in (Tresp and Molitor, 1998), the truth value of an assertion is replaced by a membership value from $[0,1]$. $\mathcal{ALC}$ assertions α in (Straccia, 2001) are made fuzzy by writing, e.g., $\langle\alpha \geq n\rangle$, thus taking a single truth value from $[0,1]$. An even more expressive theoretical description logic, Fuzzy OWL, based on OWL DL, is investigated in (Stoilos et al., 2005).

Our work might also be viewed as a modalized version of a restricted fragment of *Subjective Logic* (Jøsang, 1997; Jøsang, 2001), a probabilistic logic that can be seen as an extension of Dempster-Shafer belief theory (Wilson, 2000). Subjective Logic addresses subjective believes by requiring numerical values for *believe b*, *disbelieve d*, and *uncertainty u*, called (subjective) opinions. For each proposition, it is required that $b+d+u=1$.

The translation from modals δ to $\langle b,d,u\rangle$ is determined by the length of the confidence interval $\mu(\delta) = [l,h]$ and its starting/ending numbers, viz., $u := h-l$, $b := l$, and $d := 1-h$ (cf. Figure 2).

These definitions also address *in-the-middle* modals (cf. footnote 3). Such modals even do *not* need to be symmetrical, i.e., being around the center of the confidence interval. The definitions are clearly not applicable to the error modal ! (cf. Section 5.1.7) and it makes perfect sense to assume $u=0$ here (remember, $\mu(!) = \emptyset$), and thus bisecting the belief mass for this corner case, i.e., $b=0.5$ and $d=0.5$.

The simplification and entailment rules of the formalism (Sections 4.1 and 5) allow rule-based (forward) engines to easily implement this conservative extension of OWL. Through these rules, the formalism is *compositional* by nature and thus afflicted with all the problems, reviewers have already noted on the interplay between logic and uncertainty (Dubois and Prade, 1994). Due to the finite number of modal operators, the approach is only able to approximately compute the degree of uncertainty of new knowledge instead of giving more precise estimations, by combining the low/high numbers of the confidence intervals through min/max, multiplication, addition, etc. Contrary to other approaches, we do *not* talk about the uncertainty of *complex* propositions (conjunction, disjunction) or *sets* of beliefs, but instead focus merely on the uncertainty of *atomic* ABox propositions.

ACKNOWLEDGEMENTS

The research described in this paper has been funded by the German Federal Ministry of Education and Research (BMBF) through the project **HySociaTea** (Hybrid Social Teams for Long-Term Collaboration in Cyber-Physical Environments, grant no. 01IW14001). I have profited from discussions with my colleagues Miroslav Janíček, Bernd Kiefer, and Stefan Schulz and would like to thank the ICAART reviewers for their detailed and useful suggestions—thank you guys!

REFERENCES

Areces, C., Hoffmann, G., and Denis, A. (2010). Modal logics with counting. In *Proceedings of the 17th Workshop on Logic, Language, Information and Computation (WoLLIC)*, pages 98–109.

Bishop, B., Kiryakov, A., Ognyanoff, D., Peikov, I., Tashev, Z., and Velkov, R. (2011). OWLIM: A family of scalable semantic repositories. *Semantic Web*, 2(1):33–42.

Blackburn, P., de Rijke, M., and Venema, Y. (2001). *Modal Logic*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge.

Carroll, J. J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., and Wilkinson, K. (2004). Jena: Implementing the Semantic Web recommendations. In *Proceedings of the 13th international World Wide Web conference (WWW)*, pages 74–83.

Devlin, K. (2006). Situation theory and situation semantics. In Gabbay, D. M. and Woods, J., editors, *Handbook of the History of Logic. Volume 7*, pages 601–664. Elsevier.

Dubois, D. and Prade, H. (1994). Can we enforce full compositionality in uncertainty calculi. In *Proceedings of the 12th National Conference on Artificial Intelligence*, pages 149–154.

Fine, K. (1972). In so many possible worlds. *Notre Dame Journal of Formal Logic*, 13(4):516–520.

Golbreich, C. and Wallace, E. K. (2012). OWL 2 web ontology language new features and rationale (second edition). Technical report, W3C.

Halpern, J. Y. (1990). An analysis of first-order logics of probability. *Artificial Intelligence*, 46:311–350.

Halpern, J. Y. (2003). *Reasoning About Uncertainty*. MIT Press, Cambridge, MA.

Hayes, P. (2004). RDF semantics. Technical report, W3C.

Heinsohn, J. (1993). *ALCP – Ein hybrider Ansatz zur Modellierung von Unsicherheit in terminologischen Logiken*. PhD thesis, Universität des Saarlandes. In German.

Jaeger, M. (1994). Probabilistic reasoning in terminological logics. In *Proceedings of the 4th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 305–316.

Jøsang, A. (1997). Artificial reasoning with subjective logic. In *Proceedings of the 2nd Australian Workshop on Commonsense Reasoning*.

Jøsang, A. (2001). A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzzyness and Knowledge-Based Systems*, 9(3):279–311.

Koller, D. and Friedmann, N. (2009). *Probabilistic Graphical Models*. MIT Press.

Krengel, U. (2003). *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg, 7th edition. In German.

Krieger, H.-U. (2012). A temporal extension of the Hayes/ter Horst entailment rules and an alternative to W3C's n-ary relations. In *Proceedings of the 7th International Conference on Formal Ontology in Information Systems (FOIS)*, pages 323–336.

Krieger, H.-U. (2013). An efficient implementation of equivalence relations in OWL via rule and query rewriting. In *Proceedings of the 7th IEEE International Conference on Semantic Computing (ICSC)*, pages 260–263.

Lucas, P. J., van der Gaag, L. C., and Abu-Hanna, A. (2004). Bayesian networks in biomedicine and health-care. *Artificial Intelligence in Medicine*, 30:201–214.

Motik, B., Cuenca Grau, B., Horrocks, I., Wu, Z., Fokoue, A., and Lutz, C. (2012). OWL 2 web ontology language profiles. Technical report, W3C. W3C Recommendation 11 December 2012.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, CA, MA.

Richardson, M. and Domingos, P. (2006). Markov logic networks. *Machine Learning*, 62(1–2):107–136.

Schulz, S., Martínez-Costa, C., Karlsson, D., Cornet, R., Brochhausen, M., and Rector, A. (2014). An ontological analysis of reference in health record statements. In *Proceedings of the 8th International Conference on Formal Ontology in Information Systems (FOIS 2014)*.

Stoilos, G., Stamou, G. B., Tzouvaras, V., Pan, J. Z., and Horrocks, I. (2005). Fuzzy OWL: uncertainty and the semantic web. In *Proceedings of the OWLED '05 Workshop on OWL: Experiences and Directions*.

Straccia, U. (2001). Reasoning within fuzzy description logics. *Journal of Artificial Intelligence Research*, 14:147–176.

ter Horst, H. J. (2005). Completeness, decidability and complexity of entailment for RDF Schema and a semantic extension involving the OWL vocabulary. *Journal of Web Semantics*, 3:79–115.

Tresp, C. B. and Molitor, R. (1998). A description logic for vague knowledge. In *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI)*, pages 361–365.

Wikipedia (2015). Modal logic — Wikipedia, The Free Encyclopedia. [Online; accessed 19-June-2015].

Wilson, N. (2000). Algorithms for Dempster-Shafer theory. In *Algorithms for Uncertainty and Defeasible Reasoning*, Handbook of Defeasible Reasoning and Uncertainty Management Systems, pages 421–475. Kluwer.

×-*Protégé*
An Ontology Editor for Defining Cartesian Types to Represent n-ary Relations

Christian Willms, Hans-Ulrich Krieger, Bernd Kiefer
German Research Center for Artificial Intelligence (DFKI)
Campus D3 2, Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
christian.willms@dfki.de, krieger@dfki.de, kiefer@dfki.de

Abstract

Arbitrary n-ary relations ($n \geq 1$) can, in principle, be realized through binary relations obtained by a reification process which introduces new individuals to which the additional arguments are linked via "accessor" properties. Modern ontologies which employ standards such as RDF and OWL have mostly obeyed this restriction, but have struggled with it nevertheless. In (Krieger and Willms, 2015), we have laid the foundations for a *theory-agnostic* extension of RDFS and OWL and have implemented in the last year an extension of *Protégé*, called ×-*Protégé*, which supports the definition of Cartesian types to represent n-ary relations and relation instances. Not only do we keep the distinction between the domain and the range of an n-ary relation, but also introduce so-called *extra* arguments which can be seen as position-oriented unnamed *annotation* properties and which are accessible to entailment rules. As the direct representation of n-ary relations abolishes RDF triples, we have backed up ×-*Protégé* by the semantic repository and entailment engine *HFC* which supports tuples of arbitrary length. ×-*Protégé* is programmed in Java and is made available under the Mozilla Public License.

Keywords: ontology editor, ×-*Protégé*, Cartesian types, n-ary relations, RDF, RDFS, OWL, n-ary Description Logics.

1. Decription Logics, OWL, and RDF

Relations in description logics (DLs) are either unary (so-called *concepts* or *classes*) or binary (*roles* or *properties*) predicates (Baader et al., 2003). As the designers of OWL (Smith et al., 2004; Hitzler et al., 2012) decided to be compatible with already existing standards, such as RDF (Cyganiak et al., 2014) and RDFS (Brickley and Guha, 2014), as well as with the universal RDF data object, the *triple*,

subject predicate object

a unary relation such as $C(a)$ (class membership) becomes a binary relation via the RDF type predicate:

a rdf:type C

For very good reasons (mostly for decidability), DLs usually restrict themselves to decidable function-free two-variable subsets of first-order predicate logic. Nevertheless, people have argued ver early for relations of more than two arguments (Schmolze, 1989), some of them still retaining decidability and coming up with a better memory footprint and a better complexity for the various inference tasks (including querying) than their triple-based relatives (Krieger, 2012; Krieger, 2014). This idea conservatively extends the standard *triple-based* model towards a more general *tuple-based* approach ($n+1$ being the arity of the *predicate*):

subject predicate $object_1 \ldots object_n$

Using a standard relation-oriented notation, we often interchangeably write

$p(s, o_1, \ldots, o_n)$

Here is an example, dealing with *diachronic* relations (Sider, 2001), relation instances whose object values might change over time, but whose subject values coincide with each other. For example (quintuple representation),

peter marriedTo liz 1997 1999
peter marriedTo lisa 2000 2010

or (relation notation)

marriedTo(*peter, liz, 1997, 1999*)
marriedTo(*peter, lisa, 2000, 2010*)

which we interpret as the (time-dependent) statement that *Peter* was married to *Liz* from 1997 until 1999 and to *Lisa* from 2000–2010.

In a triple-based setting, semantically representing the same information requires a lot more effort. There already exist several approaches to achieve this (Welty and Fikes, 2006; Gangemi and Presutti, 2013; Krieger and Declerck, 2015), all coming up with at least one brand-new individual (introduced by a hidden existential quantification), acting as an *anchor* to which the object information (the range information of the relation) is bound through additional properties (a kind of *reification*). For instance, the so-called *N-ary relation encoding* (Hayes and Welty, 2006), a W3C best-practice recommendation, sticks to binary relations/triples and uses *container* objects to encode the range information (ppt1 and ppt2 being the *new* individuals):

peter marriedTo ppt1
ppt1 rdf:type nary:PersonPlusTime
ppt1 nary:value liz
ppt1 nary:starts "1997"^^xsd:gYear
ppt1 nary:ends "1999"^^xsd:gYear
peter marriedTo ppt2
ppt2 rdf:type nary:PersonPlusTime
ppt2 nary:value lisa
ppt2 nary:starts "2000"^^xsd:gYear
ppt2 nary:ends "2010"^^xsd:gYear

As we see from this small example, a quintuple is represented by five triples. The relation name is retained, however, the range of the relation changes from, say, Person to the type of the container object which we call here PersonPlusTime.

Rewriting ontologies to the *latter* representation is an unpleasant enterprise, as it requires further classes, redefines property signatures, and rewrites relation instances,

as shown by the marriedTo example above. In addition, reasoning and querying with such representations is extremely complex, expensive, and error-prone.
Unfortunately, the *former* tuple-based representation which argues for additional (temporal) arguments is **not** supported by *ontology editors* today, as it would require to deal with general n-ary relations ($n \geq 2$). ×-*Protégé* fills exactly this gap.

2. Further Motivation

×-*Protégé* supports the definition of Cartesian types, composed from standard OWL classes and XSD datatypes. Given Cartesian types and by keeping the distinction between the *domain* $\mathbb{D}$ and the *range* $\mathbb{R}$ of a binary property p, it is now possible to define $m + n$-ary relations $p \subseteq \mathbb{D}_1 \times \cdots \times \mathbb{D}_m \times \mathbb{R}_1 \times \cdots \times \mathbb{R}_n$.
The deeper reason why it is still useful to separate domain and range arguments from one another is related to the so-called *property characteristics* built into OWL, e.g., symmetry or transitivity. This ultimately allows us to generalize the corresponding entailment rules, by replacing atomic classes with Cartesian types. For instance, entailment rule rdfp4 for *transitive* properties p from (ter Horst, 2005)

$$p(x, y) \wedge p(y, z) \rightarrow p(x, z)$$

can be generalized as ($m = n = o$)

$$p(\times_{i=1}^{m} x_i, \times_{j=1}^{n} y_j) \wedge p(\times_{j=1}^{n} y_j, \times_{k=1}^{o} z_k) \\ \rightarrow p(\times_{i=1}^{m} x_i, \times_{k=1}^{o} z_k)$$

×-*Protégé* not only keeps the distinction between the *domain* and *range* arguments of a relation, but also provides further distinct *annotation*-like arguments, called *extra* arguments which have been shown useful in various situations and which are accessible to entailment rules of the above kind. Consider a binary *symmetric* property q which we would like to generalize by the concept of *valid time* (the time in which an atemporal statement is true), thus the corresponding entailment rule needs to be extended by two further temporal arguments b and e:

$$q(x, y, b, e) \rightarrow q(y, x, b, e)$$

By assuming that the temporal arguments are part of the domain and/or range of q, we are running into trouble as symmetric properties require the same number of arguments in domain and range position. Thus, we *either* need to adjust this rule, i.e.,

$$q(\underline{x, b, e}, \underline{y, b, e}) \rightarrow q(\underline{y, b, e}, \underline{x, b, e})$$

or assume that b and e have a special "status". We decided for the latter and call such information *extra arguments*. As an example, the former marriedTo relation (a symmetric relation) is of that kind, thus having the following relation signature (assuming a biography ontology with class Person):

Person ×	Person ×	xsd:gYear × xsd:gYear
domain	*range*	*2 extra arguments*

Other *non-temporal* examples of *extra arguments* might involve *space* (or *spacetime* in general), using further XSD custom types, such as point2D or point3D, in order to encode the position of a moving object over time (Keshavdas and Kruijff, 2014).

More linguistically-motivated examples include the *direct* representation of ditransitive and ergative verb frames, including adjuncts (Krieger, 2014). We will present an example of this at the end of Section 7. when defining the quaternary relation obtains. Such kinds of properties are often wrongly addressed in triple-based settings through *relation composition*, applied to the second argument of the corresponding binary relation. This does *not* work in general, but only if the original relation is *inverse functional*.
As a last example, we would like to mention the *direct* representation of *uncertain* statements in medicine or technical diagnosis in an extension of OWL (Krieger, 2016) which is far superior to various encodings described in (Schulz et al., 2014) which have accepted the boundaries of RDF triples in order to be compatible with an existing standard.

3. *Protégé*, ×-*Protégé*, and *HFC*

Protégé is a free, open source ontology editor, providing a graphical user interface to define and inspect ontologies (http://protege.stanford.edu). *Protégé* version 4 has been designed as a modular framework through the use of the OSGi framework as a plugin infrastructure (https://www.osgi.org/developer/). For this reason, ×-*Protégé* has been implemented as an EditorKitFactory plugin for *Protégé*, replacing the built-in OWL EditorKitFactory. The EditorKit is the access point for a particular type of model (in our case, a model based on n-tuples) to which a GUI has access to.

×-*Protégé* is divided into three separate components (Figure 1, large right box). The "bottom" layer is realized by *HFC* (Krieger, 2013), a bottom-up forward chainer and semantic repository implemented in Java which is comparable to popular systems such as Jena and OWLIM (http://www.dfki.de/lt/onto/hfc/). *HFC* supports RDFS and OWL reasoning à la (Hayes, 2004) and (ter Horst, 2005), but at the same time provides an expressive language for defining custom rules, involving functional and relational variables, complex tests and actions, and the replacement of triples in favour of tuples of arbitrary length. The query language of *HFC* implements a subset of SPARQL, but at the same time provides powerful custom $M{:}N$ aggregates ($M, N \geq 1$), not available in SPARQL.
The data read in by *HFC* is preprocessed and transformed into an ×-*Protégé* model. Among other things, it contains inheritance hierarchies for classes and properties which are directly used to visualize the ontology in the graphical user interface of ×-*Protégé*.
This GUI consists of several workspaces (similar to *Protégé*, version 4.3), presenting the ontology itself, the classes, the properties, and the instances. User actions result in an update of the model and *HFC*'s n-tuple database.

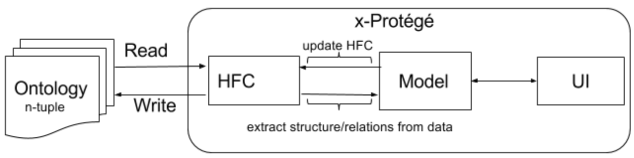


Figure 1: The three-layered structure of ×-*Protégé*.

In the next section, we will look into some of these workspaces (or tabs), assuming the marriedTo example from Sections 1. and 2.

4. Class Tab

When starting ×-*Protégé* the class hierarchy consists of a unique, most general type, called Thing+ in the GUI which subsumes every other Cartesian type and which can be formally defined as

$$\text{Thing+} := \bigsqcup_{i=1}^{k} (\text{owl:Thing} \sqcup \text{xsd:AnyType})^i$$

For a given ontology, k is fixed (finite, of course). Initially, Thing+ has two direct subtypes, viz., owl:Thing and xsd:AnyType. *HFC* already provides a set of built-in XSD subtypes, such as xsd:gYear (Gregorian Year) or xsd:int (4 Byte integers), but also defines non-standard datatypes, such as xsd:monetary. As in a pure OWL setting, owl:Thing and xsd:AnyType are incompatible, but xsd:AnyType is made available under Thing+ in order to define Cartesian types, such as xsd:gYear × xsd:gYear for the two extra arguments of the marriedTo relation (or even Person × xsd:gYear × xsd:gYear for the sexternary relation q in Section 2.). This small type hierarchy is depicted in Figure 2.

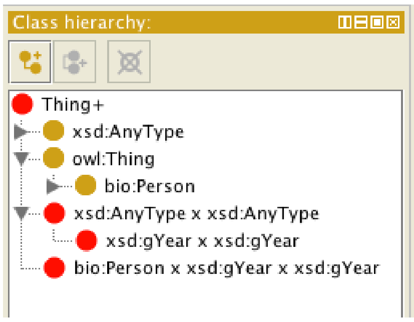


Figure 2: The class hierarchy for the marriedTo example.

Note that the non-singleton Cartesian types are highlighted using red colour and that xsd:gYear × xsd:gYear is correctly classified as a subclass of the Cartesian type xsd:AnyType × xsd:AnyType.

5. Property Tab

As in OWL, we distinguish between the property characteristics owl:DatatypeProperty and owl:ObjectProperty. We group these two classes under the super-property MixedProperty, as we do allow for further "mixed" property characteristics; e.g., properties which are instantiated with an XSD atom in first place or properties with Cartesian domain and range types which are a mixture of OWL classes and XSD types (and thus are neither datatype nor object properties). Since the quaternary relation marriedTo (binary relation plus two extra args) maps URIs onto URIs, it is classified as an object property (remember, the extra args neither belong to the domain nor range of a property). However, the ternary relation hasAge (binary relation plus one extra args) is a datatype property as it maps URIs onto XSD ints (the extra arg is the *transaction time*, the time when the birthdate was entered to *HFC*); cf. Figure 3.

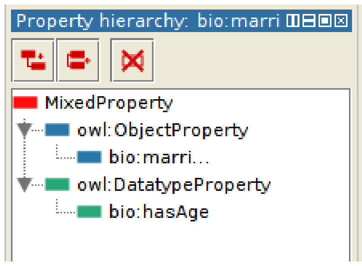


Figure 3: The property hierarchy for the marriedTo and hasAge relations.

Figure 4: The property signature for the marriedTo relation.

When defining a new property, a user is required to choose the right Cartesian types to complete the property signature. This is displayed in Figure 4 for the marriedTo relation. Depending on the kind of property, an ontology engineer is even allowed to associate further property characteristics with a property under definition; see Figure 5.

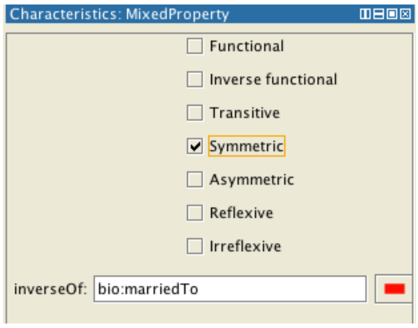


Figure 5: Further potential property characteristics for the marriedTo relation.

6. Instance Tab

We complete the overview of the workspace tabs by coming back to *Peter* and his relation to *Liz* and *Lisa* (cf. Section 1.). From the instance tab, we learn about his two marriages and that he is currently 53 years old (see Figure 6).
The symmetry of the marriedTo relation (see Figure 5) further guarantees that *Peter* is listed in the instance tabs of *Liz* and *Liza* as well.

7. N-Tuples & I/O Formats

As ×-*Protégé* allows us to deviate from pure binary relations, certain adjustments to the *N-triples* format (Carothers and Seaborne, 2014) are necessary, especially as extra arguments need to be represented. Assume a quaternary relation obtains between a person and a degree obtained from an educational organization at a specific time:

$$\text{obtains} \subseteq \underbrace{\text{Person}}_{\mathbb{D}} \times \underbrace{\text{Degree} \times \text{School}}_{\mathbb{R}_1 \times \mathbb{R}_2} \times \underbrace{\text{xsd:date}}_{\mathbb{A}}$$

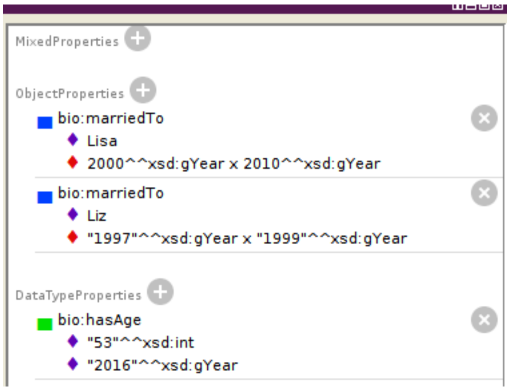


Figure 6: Facts about *Peter*.

In order to let the system know of how many arguments the domain, the range, and the extra part of a relation is composed of, we add further length-related information (infix notation):

```
obtains rdfs:domain Person
obtains rdfs:range Degree School
obtains nary:extra xsd:date
obtains nary:domainArity "1"^^xsd:int
obtains nary:rangeArity "2"^^xsd:int
obtains nary:extraArity "1"^^xsd:int
```

Notice that the rdfs:range keyword directly above is followed by *two* classes: Degree and School ($= \mathbb{R}_1 \times \mathbb{R}_2$). Not only is this kind of representation used in the RBox of an ontology, but also in the TBox, e.g.

```
Degree School rdfs:subClassOf owl:Thing owl:Thing
```

as

Degree $\times$ School $\sqsubseteq \top \times \top$

is the case. ABox information is also affected by this style of representation, as, for instance

```
peter obtains phd stanford "1985"^^xsd:date
```

Besides providing such an (asymmetric) *infix* representation, $\times$-*Protégé* let the user decide whether a *prefix* representation is more appropriate for him/her. So, for instance, the last ABox statement above would then become

```
obtains peter phd stanford "1985"^^xsd:date
```

We finally like to stress the fact that once one decided to go for a direct representation of additional arguments and reason upon them, queries and rules will usually intermix tuples of different length. For example, in a *valid time* approach *universal* information from the TBox and RBox of an ontology is encoded as triples, whereas *assertional* knowledge will be represented as quintuples (Krieger, 2012); see *HFC* rule at the end of Section 8.

8. Future Work

Since $\times$-*Protégé* already uses functionality from *HFC* (see Section 3.), we would like to add further *query* and *rule definition* tabs to the next major version of $\times$-*Protégé* to support the construction of *HFC* queries and rules (see the two examples below).

The query support in $\times$-*Protégé* will ease the definition of SPARQL-like queries in *HFC* over n-tuples, using keywords such as SELECT, SELECTALL (for the *multiply-out* mode in *HFC* in case equivalence class reduction is enabled), DISTINCT, WHERE, FILTER, and AGGREGATE. Depending on the property signatures, $\times$-*Protégé* will then alarm a user if too less, too many, or wrong arguments have been specified in WHERE clauses, FILTER tests, or AGGREGATE functions. This helps to simply the construction of a query such as

```
SELECT DISTINCT ?partner
WHERE peter marriedTo ?partner ?start ?end
FILTER GreaterEqual ?start "1998"^^xsd:gYear &
         LessEqual ?end "2005"^^xsd:gYear
AGGREGATE ?noOfPartners = Count ?partner
```

which computes how many times *Peter* was married to distinct women between 1998 and 2005. The results of such queries (viz., tables) will also be displayed in this tab.

The rule support will provide means to define, maintain, and extend RDFS, OWL, and custom rule sets. Again, as is the case for queries, clauses, @test, and @action sections of rules in *HFC* will benefit from checking for the right number of arguments. For instance, the valid time extension of the entailment rule for *transitive* properties (ter Horst, 2005) in *HFC* looks as follows (Krieger, 2012):

```
?p rdf:type owl:TransitiveProperty     // triple
?x ?p ?y ?start1 ?end1                 // quintuple
?y ?p ?z ?start2 ?end2
→
?x ?p ?z ?start ?end
@test                                  // 3 LHS tests
?x != ?y
?y != ?z
IntersectionNotEmpty ?start1 ?end1 ?start2 ?end2
@action                                // 2 RHS actions
?start = Max2 ?start1 ?start2          // new RHS variable
?end = Min2 ?end1 ?end2                // new RHS variable
```

In both cases, we would also like to provide a *completion* mechanism for properties and URIs, as well as for external tests (see @test above) and value-returning functions (see @action above), an extremely useful functionality known from programming environments.

Our ultimate goal is thus to offer $\times$-*Protégé* as a front-end GUI for ontology-based systems, based on *HFC*.

9. Download

$\times$-*Protégé* version 1.0 as of Monday Feb 15, 2016 can be downloaded from https://bitbucket.org/cwillms/x-protege/downloads/ and is made available under the Mozilla Public License. Here, you will also find a preliminary version of the user guide.

10. Acknowledgements

The research described in this paper has been partially financed by the European project PAL (Personal Assistant for healthy Lifestyle) under Grant agreement no. 643783-RIA Horizon 2020. This work was conducted using the *Protégé* resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health. We would like to thank the three reviewers for their detailed and useful suggestions.

11. Bibliographical References

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., and Patel-Schneider, P. (2003). *The Description Logic Handbook*. Cambridge University Press, Cambridge.

Brickley, D. and Guha, R. (2014). RDF Schema 1.1. Technical report, W3C.

Carothers, G. and Seaborne, A. (2014). RDF 1.1 N-Triples. a line-based syntax for an RDF graph. Technical report, W3C.

Cyganiak, R., Wood, D., and Lanthaler, M. (2014). RDF 1.1 concepts and abstract syntax. Technical report, W3C.

Gangemi, A. and Presutti, V. (2013). A multi-dimensional comparison of ontology design patterns for representing n-ary relations. In *39th International Conference on Current Trends in Theory and Practice of Computer Science*, pages 86–105.

Hayes, P. and Welty, C. (2006). Defining N-ary relations on the Semantic Web. Technical report, W3C.

Hayes, P. (2004). RDF semantics. Technical report, W3C.

Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F., and Rudolph, S. (2012). OWL 2 web ontology language primer (second edition). Technical report, W3C.

Keshavdas, S. and Kruijff, G.-J. (2014). Functional mapping for human-robot cooperative exploration. *International Journal of Computer and Applications*, 36(1).

Krieger, H.-U. and Declerck, T. (2015). An OWL ontology for biographical knowledge. representing time-dependent factual knowledge. In *Proceedings of the Workshop on Biographical Data in a Digital World*.

Krieger, H.-U. and Willms, C. (2015). Extending OWL ontologies by Cartesian types to represent N-ary relations in natural language. In *Proceedings of the IWCS Workshop on Language and Ontologies*.

Krieger, H.-U. (2012). A temporal extension of the Hayes/ter Horst entailment rules and an alternative to W3C's n-ary relations. In *Proceedings of the 7th International Conference on Formal Ontology in Information Systems (FOIS)*, pages 323–336.

Krieger, H.-U. (2013). An efficient implementation of equivalence relations in OWL via rule and query rewriting. In *Proceedings of the 7th IEEE International Conference on Semantic Computing (ICSC)*, pages 260–263.

Krieger, H.-U. (2014). A detailed comparison of seven approaches for the annotation of time-dependent factual knowledge in RDF and OWL. In *Proceedings of the 10th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA)*.

Krieger, H.-U. (2016). Capturing graded knowledge and uncertainty in a modalized fragment of OWL. In *Proceedings of the 8th International Conference on Agents and Artificial Intelligence (ICAART)*, pages 19–30.

Schmolze, J. G. (1989). Terminological knowledge representation systems supporting n-ary terms. In *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 432–443.

Schulz, S., Martínez-Costa, C., Karlsson, D., Cornet, R., Brochhausen, M., and Rector, A. (2014). An ontological analysis of reference in health record statements. In *Proceedings of the 8th International Conference on Formal Ontology in Information Systems (FOIS 2014)*.

Sider, T. (2001). *Four Dimensionalism. An Ontology of Persistence and Time*. Oxford University Press.

Smith, M. K., Welty, C., and McGuinness, D. L. (2004). OWL Web Ontology Language Guide. Technical report, W3C.

ter Horst, H. J. (2005). Completeness, decidability and complexity of entailment for RDF Schema and a semantic extension involving the OWL vocabulary. *Journal of Web Semantics*, 3:79–115.

Welty, C. and Fikes, R. (2006). A reusable ontology for fluents in OWL. In *Proceedings of 4th FOIS*, pages 226–236.

A Modal Representation of Graded Medical Statements

Hans-Ulrich Krieger[1] and Stefan Schulz[2]

[1] German Research Center for Artificial Intelligence (DFKI) `krieger@dfki.de`
[2] Institute of Medical Informatics, Medical University of Graz
`stefan.schulz@medunigraz.at`

Abstract. Medical natural language statements uttered by physicians are usually *graded*, i.e., are associated with a degree of uncertainty about the validity of a medical assessment. This uncertainty is often expressed through specific *verbs*, *adverbs*, or *adjectives* in natural language. In this paper, we look into a *representation* of such graded statements by presenting a simple non-normal *modal logic* which comes with a set of modal operators, directly associated with the words indicating the uncertainty and interpreted through *confidence intervals* in the model theory. We complement the model theory by a set of RDFS-/OWL 2 RL-like entailment (*if-then*) rules, acting on the syntactic representation of modalized statements. Our interest in such a formalization is related to the use of OWL as the *de facto* standard in (medical) ontologies today and its weakness to represent and reason about assertional knowledge that is uncertain or that changes over time. The approach is not restricted to medical statements, but is applicable to other graded statements as well.

1 Introduction & Background

Medical natural language statements uttered by physicians or other health professionals and found in medical examination letters are usually *graded*, i.e., are associated with a degree of uncertainty about the validity of a medical assessment. This uncertainty is often expressed through specific verbs, adverbs, or adjectives in natural language (which we will call *gradation words*). E.g., *Dr. X suspects that Y suffers from Hepatitis* or *The patient probably has Hepatitis* or *(The diagnosis of) Hepatitis is confirmed*.

In this paper, we look into a representation of such graded statements by presenting a simple non-standard modal logic which comes with a small set of partially-ordered modal operators, directly associated with the words indicating the uncertainty and interpreted through confidence intervals in the model theory. The approach currently only addresses modalized propositional formulae in negation normal form which can be seen as a canonical representation of natural language sentences of the above form (a kind of a *controlled natural language*). Our interest in such a formalization is related to the use of OWL in our projects as the *de facto* standard for (medical) ontologies today and its weakness to represent and reason about assertional knowledge that is uncertain [15] or that

changes over time [12]. There are two principled ways to address such a restriction: *either* by sticking with the existing formalism (viz., OWL) and trying to find an encoding that still enables some useful forms of reasoning [15]; *or* by deviating from a defined standard in order to arrive at an easier, intuitive, and less error-prone representation [12].
Here, we follow the latter avenue, but employ and extend the standard entailment rules from [6] and [18] for positive binary relation instances in RDFS and OWL towards modalized n-ary relation instances, including negation. These entailment rules talk about, e.g., subsumption, class membership, or transitivity, and have been found useful in many applications. The proposed solution has been implemented in *HFC* [13], a forward chaining engine that builds Herbrand models which are compatible with the open-world view underlying OWL. The approach presented in this paper is clearly not restricted to medical statements, but is applicable to other graded statements as well (including *trust*), e.g., technical diagnosis (*The engine is probably overheated*) or more general in everyday conversation (*I'm pretty sure that X has signed a contract with Y*) which can be seen as the common case (contrary to true *universal* statements).

2 Graded Medical Statements: OWL vs. Modalized Representation

We note here that our initial modal operators were inspired by the *qualitative information parts* of diagnostic statements from [15] shown in Figure 1, but we might have chosen other operators, capturing the meaning of the gradation words used in the examples at the beginning of Section 1 (e.g., *probably*).

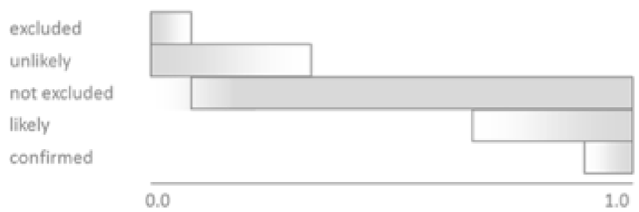


Fig. 1. Vague schematic mappings of the qualitative information parts *excluded* (E), *unlikely* (U), *not excluded* (N), *likely* (L), and *confirmed* (C) to *confidence intervals*, as used in this paper. Figure taken from [15].

These qualitative parts were used in statements about, e.g., liver inflammation with varying levels of detail. From this, we want to infer that, e.g., **if** *Hepatitis is confirmed* **then** *Hepatitis is likely* but **not** *Hepatitis is unlikely*. And **if** *Viral Hepatitis B is confirmed*, **then** both *Viral Hepatitis is confirmed* **and** *Hepatitis is confirmed* (generalization). Things "turn around" when we look at the adjectival modifiers *excluded* and *unlikely*: **if** *Hepatitis is excluded* **then** *Hepatitis is unlikely*, but **not** *Hepatitis is not excluded*. Furthermore, **if** *Hepatitis is excluded*,

then both *Viral Hepatitis is excluded* **and** *Viral Hepatitis B is excluded* (specialization). The set of *plausible* entailments for this kind of graded reasoning is depicted in Figure 2.

		Being said to have hepatitis (H) / viral hepatitis (vH) / viral hepatitis B (vHB) is...														
Precondition:		confirmed			likely			not excluded			unlikely			excluded		
Entailment:		H	vH	vHB	H	vH	vHB	H	vH	vHB	H	vH	vHB	H	vH	vHB
confirmed	H	x	x	x												
	vH		x	x												
	vHB			x												
likely	H	x	x	x	x	x	x									
	vH		x	x		x	x									
	vHB			x			x									
not excluded	H	x	x	x	x	x	x	x	x	x						
	vH		x	x		x	x		x	x						
	vHB			x			x			x						
unlikely	H										x			x		
	vH										x	x		x	x	
	vHB										x	x	x	x	x	x
excluded	H													x		
	vH													x	x	
	vHB													x	x	x

Fig. 2. Statements about liver inflammation with varying levels of detail: *Viral Hepatitis B* (vHB) implies *Viral Hepatitis* (vH) which implies *Hepatitis* (H). The matrix depicts entailments considered plausible, based on the inferences that follow from Figure 1. Hepatitis and its subclasses can be easily replaced by other medical situations/diseases. Figure taken from [15].

[15] consider five encodings (one outside the expressivity of OWL), from which only two were able to fully reproduce the inferences from Figure 2. Let us quickly look on approach 1, called *existential restriction*, before we informally present its modal counterpart (we will use abstract description logic syntax here [2]):

$$\mathsf{HepatitisSituation} \equiv \mathsf{ClinicalSituation} \sqcap \exists \mathsf{hasCondition.Hepatitis}$$

% *Hepatitis subclass hierarchy*

$$\mathsf{ViralHepatitisB} \sqsubseteq \mathsf{ViralHepatitis} \sqsubseteq \mathsf{Hepatitis}$$

% *vagueness via two subclass hierarchies*

$$\mathsf{IsConfirmed} \sqsubseteq \mathsf{IsLikely} \sqsubseteq \mathsf{IsNotExcluded} \qquad \mathsf{IsExcluded} \sqsubseteq \mathsf{IsUnlikely}$$

% *a diagnostic statement about Hepatitis*

$$\mathsf{BeingSaidToHaveHepatitisIsConfirmed} \equiv \mathsf{DiagnosticStatement} \sqcap \forall \mathsf{hasCertainty.IsConfirmed} \sqcap \exists \mathsf{isAboutSituation.HepatitisSituation}$$

Standard OWL reasoning under this representation then ensures that, for instance,

$$\mathsf{BeingSaidToHaveHepatitsIsConfirmed} \sqsubseteq \mathsf{BeingSaidToHaveHepatitisIsLikely}$$

is the case, exactly one of the plausible inferences from Figure 2.
The encodings in [15] were quite cumbersome as the primary interest was to stay within the limits of the underlying calculus (OWL). Besides coming up with complex encodings, only minor forms of reasoning were possible, viz., subsumption reasoning. These disadvantages are a result of two conscious decisions:

OWL only provides unary and binary relations (concepts and roles) and comes up with a (mostly) fixed set of entailment/tableaux rules.
In our approach, however, the *qualitative information parts* from Figure 1 are first class citizens of the object language (the modal operators) and *diagnostic statements* from the Hepatitis use case are expressed through the binary property $\mathsf{suffersForm}$ between p (patients, people) and d (diseases, diagnoses). The plausible inferences are then simply a *byproduct* of the *instantiation* of the entailment rule schemas (G) from Section 5.1, and (S1) and (S0) from Section 5.2 for property $\mathsf{suffersForm}$ (the rule variables are universally quantified; $\top$ = *universal truth*; C = *confirmed*; L = *likely*), e.g.,

(S1) $\top\mathsf{ViralHepatitisB}(d) \wedge \mathsf{ViralHepatitisB} \sqsubseteq \mathsf{ViralHepatitis} \rightarrow \top\mathsf{ViralHepatitis}(d)$
(G) $C\mathsf{suffersFrom}(p,d) \rightarrow L\mathsf{suffersFrom}(p,d)$

Two things are worth to be mentioned here. *Firstly*, not only OWL-like properties (binary relations) can be graded, such as $C\mathsf{suffersFrom}(p,d)$ (= *it is confirmed that p suffers from d*), but also class membership (unary relations), e.g., $C\mathsf{ViralHepatitisB}(d)$ (= *it is confirmed that d is Viral Hepatitis B*). However, as the original OWL example above is unable to make use of any modals, we employ a special modal $\top$ here:$\top\mathsf{ViralHepatitisB}(d)$. *Secondly*, modal operators are only applied to assertional knowledge, involving individuals (the ABox in OWL)—neither axioms about classes (TBox) nor properties (RBox) are being affected by modals, as they are supposed to express universal truth.

3 Confidence of Statements and Confidence Intervals

We address the *confidence* of an asserted medical statement [15] through *graded* modalities applied to propositional formulae: E (*excluded*), U (*unlikely*), N (*not excluded*), L (*likely*), and C (*confirmed*). For various (technical) reasons, we add a *wildcard* modality ? (*unknown*), a complementary *failure* modality ! (*error*), plus two further modalities to syntactically state definite truth and falsity: $\top$ (*true*) and $\bot$ (*false*). Let $\triangle$ now denotes the set of all modalities:

$$\triangle = \{?, !, \top, \bot, E, U, N, L, C\}$$

A *measure function*

$$\mu : \triangle \mapsto [0,1] \times [0,1]$$

is a mapping which returns the associated *confidence interval* $[l,h]$ for a modality from $\triangle$ $(l \leq h)$. We presuppose that

- $\mu(?) = [0,1]$
- $\mu(!) = \emptyset$[3]
- $\mu(\top) = [1,1]$
- $\mu(\bot) = [0,0]$

In addition, we define two disjoint subsets of $\triangle$, called

- $\mathsf{1} = \{\top, C, L, N\}$
- $\mathsf{0} = \{\bot, E, U\}$

[3] Recall that an interval is a set of real numbers, together with a total ordering relation (e.g., $\leq$) over the elements, thus $\emptyset$ is a perfect, although degraded interval.

and again make a presupposition: the confidence intervals for modals from 1 *end* in 1, whereas the confidence intervals for 0 modals always *start* with 0. It is worth noting that we do *not* make use of μ in the syntax of the modal language (for which we employ the modalities from $\triangle$), but in the semantics when dealing with the satisfaction relation of the model theory (see Section 4).

We have talked about *confidence intervals* now several times without saying what we actually mean by this. Suppose that a physician says that it is *confirmed* (= C) that patient p suffers from disease d, given a set of recognized symptoms $S = \{s_1, \ldots, s_k\}$: $C\,\mathit{suffersFrom}(p, d)$.

Assuming that a different patient p' shows the same symptoms S (and only S, and perhaps further symptoms which are, however, *independent* from S), we would assume that the same doctor would diagnose $C\,\mathit{suffersFrom}(p', d)$.

Even an other, but similar trained physician is supposed to grade the two patients *similarly*. This similarity which originates from patients showing the same symptoms and from physicians being taught at the same medical school is addressed by confidence *intervals* and not through a *single* (posterior) probability, as there are still variations in diagnostic capacity and daily mental state of the physician. By using intervals (instead of single values), we can usually reach a consensus among people upon the *meaning* of gradation words, even though the low/high values of the confidence interval for, e.g., *confirmed* might depend on the context.

Being a bit more theoretic, we define a *confidence interval* as follows. Assume a *Bernoulli experiment* [11] that involves a large set of n patients P sharing the same symptoms S. W.r.t. our example, we would like to know whether $\mathit{suffersFrom}(p, d)$ or $\neg\mathit{suffersFrom}(p, d)$ is the case for every patient $p \in P$, sharing S. Given a Bernoulli trials sequence $\boldsymbol{X} = \langle x_1, \ldots, x_n \rangle$ with indicator random variables $x_i \in \{0, 1\}$ for a patient sequence $\langle p_1, \ldots, p_n \rangle$, we can *approximate* the *expected value* E for *suffersFrom* being *true*, given disease d and background symptoms S by the *arithmetic mean* A:

$$\mathrm{E}[\boldsymbol{X}] \approx \mathrm{A}[\boldsymbol{X}] = \frac{\sum_{i=1}^{n} x_i}{n}$$

Due to the *law of large numbers*, we expect that if the number of elements in a trials sequence goes to infinity, the arithmetic mean will coincide with the expected value:

$$\mathrm{E}[\boldsymbol{X}] = \lim_{n \to \infty} \frac{\sum_{i=1}^{n} x_i}{n}$$

Clearly, the arithmetic mean for each new *finite* trials sequence is different, but we can try to *locate* the expected value within an interval around the arithmetic mean:

$$\mathrm{E}[\boldsymbol{X}] \in [\mathrm{A}[\boldsymbol{X}] - \epsilon_1, \mathrm{A}[\boldsymbol{X}] + \epsilon_2]$$

For the moment, we assume $\epsilon_1 = \epsilon_2$, so that $\mathrm{A}[\boldsymbol{X}]$ is in the center of this interval which we will call from now on *confidence interval*.

Coming back to our example and assuming $\mu(C) = [0.9, 1]$, $C\,\mathit{suffersFrom}(p, d)$ can be read as being true in 95% of all cases *known* to the physician, involving patients p potentially having disease d and sharing the same prior symptoms (evidence) $s_1, \ldots, s_k$:

$$\frac{\sum_{p \in P} \text{Prob}(\mathit{suffersFrom}(p,d)|s_1,\ldots,s_k)}{n} \approx 0.95$$

The variance of $\pm 5\%$ is related to varying diagnostic capabilities between (comparative) physicians, daily mental form, undiscovered important symptoms or examinations which have not been carried out (e.g., lab values), or perhaps even the physical stature of the patient which unconsciously affects the final diagnosis, etc, as elaborated above. Thus the individual modals from $\triangle$ express (via μ) different forms of the physician's confidence, depending on the set of already acquired symptoms as (potential) explanations for a specific disease.

4 Model Theory and Negation Normal Form

Let $\mathcal{C}$ denote the set of constants that serve as the arguments of a relation instance. In order to define basic n-ary propositional formulae (ground atoms, propositional letters), let $p(\boldsymbol{c})$ abbreviates $p(c_1, \ldots, c_n)$, for some $c_1, \ldots, c_n \in \mathcal{C}$, given $\mathit{length}(\boldsymbol{c}) = n$. In case the number of arguments do not matter, we sometimes simply write p, instead of, e.g., $p(c, d)$ or $p(\boldsymbol{c})$. As before, we assume $\triangle = \{?, !, \top, \bot, E, U, N, L, C\}$. We inductively define the set of *well-formed formulae* ϕ of our modal language as follows:

$$\phi ::= p(\boldsymbol{c}) \mid \neg\phi \mid \phi \wedge \phi' \mid \phi \vee \phi' \mid \triangle\phi$$

4.1 Simplification and Normal Form

We now syntactically *simplify* the set of well-formed formulae ϕ by restricting the uses of *negation* and *modalities* to the level of propositional letters p and call the resulting language Λ:

$$\pi ::= p(\boldsymbol{c}) \mid \neg p(\boldsymbol{c})$$

$$\phi ::= \pi \mid \triangle\pi \mid \phi \wedge \phi' \mid \phi \vee \phi'$$

To do so, we need the notion of a *complement* modal δ^{C} for every $\delta \in \triangle$, where

$$\mu(\delta^{\mathsf{C}}) := \mu(\delta)^{\mathsf{C}} = \mu(?) \setminus \mu(\delta) = [0,1] \setminus \mu(\delta)$$

I.e., $\mu(\delta^{\mathsf{C}})$ is defined as the complementary interval of $\mu(\delta)$ (within the bounds of $[0,1]$, of course). For example, E and N (*excluded, not excluded*) or ? and ! (*unknown, error*) are already existing complementary modals. We also require *mirror* modals δ^{M} for every $\delta \in \triangle$ whose confidence interval $\mu(\delta^{\mathsf{M}})$ is derived by "mirroring" $\mu(\delta)$ to the opposite site of the confidence interval, either to the left or to the right:

$$\textbf{if } \mu(\delta) = [l,h] \textbf{ then } \mu(\delta^{\mathsf{M}}) := [1-h, 1-l]$$

For example, E and C (*excluded, confirmed*) or $\top$ and $\bot$ (*top, bottom*) are mirror modals. In order to transform ϕ into its *negation normal form*, we need to apply simplification rules a finite number of times (until rules are no longer applicable). We depict those rules by using the $\vdash$ relation, read as *formula* $\vdash$ *simplified formula*:

1. $?\phi \vdash \epsilon$ % $?\phi$ *is not informative at all, but its existence should alarm us*
2. $\neg\neg\phi \vdash \phi$
3. $\neg(\phi \wedge \phi') \vdash \neg\phi \vee \neg\phi'$
4. $\neg(\phi \vee \phi') \vdash \neg\phi \wedge \neg\phi'$
5. $\neg\triangle\phi \vdash \triangle^{\mathsf{C}}\phi$ (example: $\neg E\phi = N\phi$)
6. $\triangle\neg\phi \vdash \triangle^{\mathsf{M}}\phi$ (example: $E\neg\phi = C\phi$)

Clearly, the mirror modals δ^{M} are not necessary as long as we explicitly allow for negated statements, and thus case 6 can, in principle, be dropped.
What is the result of simplifying $\triangle(\phi \wedge \phi')$ and $\triangle(\phi \vee \phi')$? Let us start with the former case and consider as an example the statement about an engine that *a mechanical failure m and an electrical failure e is confirmed*: $C(m \wedge e)$. It seems plausible to simplify this expression to $Cm \wedge Ce$. Commonsense tells us furthermore that neither Em nor Ee is compatible with this description.
Now consider the "opposite" statement $E(m \wedge e)$ which must *not* be rewritten to $Em \wedge Ee$, as *either* Cm *or* Ce is well *compatible* with $E(m \wedge e)$. Instead, we rewrite this kind of "negated" statement as $Em \vee Ee$, and this works fine with either Cm or Ce.
In order to address the other modal operators, we generalize these plausible inferences by making a distinction between $\mathsf{0}$ and $\mathsf{1}$ modals (see Section 3):

7a. $\mathsf{0}(\phi \wedge \phi') \vdash \mathsf{0}\phi \vee \mathsf{0}\phi'$
7b. $\mathsf{1}(\phi \wedge \phi') \vdash \mathsf{1}\phi \wedge \mathsf{1}\phi'$

Now let us consider disjunction inside the scope of a modal operator. As we do allow for the full set of Boolean operators, we are allowed to deduce

8. $\triangle(\phi \vee \phi') \vdash \triangle(\neg(\neg(\phi \vee \phi'))) \vdash \triangle(\neg(\neg\phi \wedge \neg\phi')) \vdash \triangle^{\mathsf{M}}(\neg\phi \wedge \neg\phi')$

This is, again, a conjunction, so we apply schemas 7a and 7b, giving us

8a. $\mathsf{0}(\phi \vee \phi') \vdash \mathsf{0}^{\mathsf{M}}(\neg\phi \wedge \neg\phi') \vdash \mathsf{1}(\neg\phi \wedge \neg\phi') \vdash \mathsf{1}\neg\phi \wedge \mathsf{1}\neg\phi' \vdash \mathsf{1}^{\mathsf{M}}\phi \wedge \mathsf{1}^{\mathsf{M}}\phi' \vdash \mathsf{0}\phi \wedge \mathsf{0}\phi'$
8b. $\mathsf{1}(\phi \vee \phi') \vdash \mathsf{1}^{\mathsf{M}}(\neg\phi \wedge \neg\phi') \vdash \mathsf{0}(\neg\phi \wedge \neg\phi') \vdash \mathsf{0}\neg\phi \vee \mathsf{0}\neg\phi' \vdash \mathsf{0}^{\mathsf{M}}\phi \vee \mathsf{0}^{\mathsf{M}}\phi' \vdash \mathsf{1}\phi \vee \mathsf{1}\phi'$

Note how the modals from $\mathsf{0}$ in 7a and 8a act as a kind of *negation* to turn the logical operators into their counterparts, similar to de Morgan's law.

4.2 Model Theory

In the following, we extend the standard definition of modal (Kripke) frames and models [3] for the *graded* modal operators from $\triangle$ by employing the measure function μ and focussing on the minimal definition for ϕ in Λ. A *frame* $\mathcal{F}$ for the probabilistic modal language Λ is a pair

$$\mathcal{F} = \langle \mathcal{W}, \triangle \rangle$$

where $\mathcal{W}$ is a non-empty set of *worlds* (or *situations, states, points, vertices*) and $\triangle$ a family of binary relations over $\mathcal{W} \times \mathcal{W}$, called *accessibility relations*. Note that we have overloaded $\triangle$ (and each $\delta \in \triangle$) in that it refers to the modals used in the syntax of Λ, but also to depict the binary relations, connecting worlds.

A *model* $\mathcal{M}$ for the probabilistic modal language Λ is a triple

$$\mathcal{M} = \langle \mathcal{F}, \mathcal{V}, \mu \rangle$$

such that $\mathcal{F}$ is a *frame*, $\mathcal{V}$ a *valuation*, assigning each proposition ϕ a subset of $\mathcal{W}$, viz., the set of worlds in which ϕ holds, and μ a mapping, returning the confidence interval for a given modality from $\triangle$. Note that we only require a definition for μ in $\mathcal{M}$ (the model, but *not* in the frame), as $\mathcal{F}$ represent the relational structure without interpreting the edge labeling (the modal names) of the graph.
The *satisfaction relation* $\models$, given a model $\mathcal{M}$ and a specific world w is inductively defined over the set of well-formed formulae of Λ in *negation normal form* (remember $\pi ::= p(\boldsymbol{c}) \mid \neg p(\boldsymbol{c})$):

1. $\mathcal{M}, w \models p(\boldsymbol{c})$ **iff** $w \in \mathcal{V}(p(\boldsymbol{c}))$ **and** $w \notin \mathcal{V}(\neg p(\boldsymbol{c}))$
2. $\mathcal{M}, w \models \neg p(\boldsymbol{c})$ **iff** $w \in \mathcal{V}(\neg p(\boldsymbol{c}))$ **and** $w \notin \mathcal{V}(p(\boldsymbol{c}))$
3. $\mathcal{M}, w \models \phi \wedge \phi'$ **iff** $\mathcal{M}, w \models \phi$ **and** $\mathcal{M}, w \models \phi'$
4. $\mathcal{M}, w \models \phi \vee \phi'$ **iff** $\mathcal{M}, w \models \phi$ **or** $\mathcal{M}, w \models \phi'$
5. **for all** $\delta \in \triangle$: $\mathcal{M}, w \models \delta\pi$ **iff** $\frac{\#\{u \mid (w,u) \in \delta \textbf{ and } \mathcal{M}, u \models \pi\}}{\#\{u \mid (w,u) \in \delta' \textbf{ and } \delta' \in \triangle\}} \in \mu(\delta)$

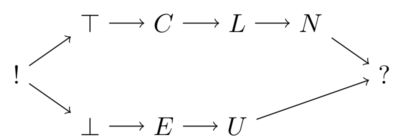
The last case of the satisfaction relation addresses the modals: for a world w, we look for the successor states u that are directly reachable via δ and in which π holds, and divide the number of such states by the number of all worlds that are directly reachable from w. This number between 0 and 1 must lie in the confidence interval $\mu(\delta)$ of δ in order to satisfy $\delta\pi$, given $\mathcal{M}, w$.
It is worth noting that the satisfaction relation above differs in its handling of $\mathcal{M}, w \models \neg p(\boldsymbol{c})$, as negation is *not* interpreted through the *absence* of $p(\boldsymbol{c})$ ($\mathcal{M}, w \not\models p(\boldsymbol{c})$), but through the *existence* of $\neg p(\boldsymbol{c})$. This treatment addresses the *open-world* nature in OWL and the evolvement of a (medical) domain over time.
We also note that the definition of the satisfaction relation for modalities (last clause) is related to the *possibility operators* $M_k \cdot$ ($= \Diamond^{\geq k} \cdot$; $k \in \mathbb{N}$) [4] and *counting modalities* $\cdot \geq n$ [1], used in modal logic characterizations of *description logics* with *cardinality* restrictions.

4.3 Well-Behaved Frames

As we will see later, it is handy to assume that the graded modals are arranged in a kind of hierarchy—the more we move "upwards" in the hierarchy, the more a statement in the scope of a modal becomes *uncertain*. In order to address this, we slightly extend the notion of a *frame* by a third component $\preceq \subseteq \triangle \times \triangle$, a partial order between modalities:

$$\mathcal{F} = \langle \mathcal{W}, \triangle, \preceq \rangle$$

Let us consider the following modal hierarchy that we build from the set $\triangle$ of already introduced modals:



This graphical representation is just a compact way to specify a set of 33 binary relation instances over $\triangle$, such as, e.g., $\top \preceq \top$, $\top \preceq N$, $C \preceq N$, $\bot \preceq ?$, or $! \preceq ?$. The above mentioned form of uncertainty is expressed by the measure function μ in that the associated confidence intervals become larger:

$$\textbf{if } \delta \preceq \delta' \textbf{ then } \mu(\delta) \subseteq \mu(\delta')$$

In order to arrive at a proper and intuitive model-theoretic semantics which mirrors intuitions such as **if** ϕ *is confirmed (*$C\phi$*)* **then** ϕ *is likely (*$L\phi$*)*, we will focus here on *well-behaved* frames $\mathcal{F}$ which enforce the existence of edges in $\mathcal{W}$, given $\preceq$ and $\delta, \delta^{\uparrow} \in \triangle$:

$$\textbf{if } (w, u) \in \delta \textbf{ and } \delta \preceq \delta^{\uparrow} \textbf{ then } (w, u) \in \delta^{\uparrow}$$

However, by imposing this constraint, we also need to adapt the last case of the satisfiability relation:

5. **for all** $\delta \in \triangle$: $\mathcal{M}, w \models \delta\pi$ **iff** $\frac{\#\{u \mid (w,u) \in \delta^{\uparrow}, \delta \preceq \delta^{\uparrow}, \textbf{ and } \mathcal{M}, u \models \pi\}}{\#\{u \mid (w,u) \in \delta' \textbf{ and } \delta' \in \triangle\}} \in \mu(\delta)$

Not only are we scanning for edges (w, u) labeled with δ and for successor states u of w in which π holds in the denominator (original definition), but also take into account edges marked with more general modals $\delta^{\uparrow}$, s.t. $\delta^{\uparrow} \succeq \delta$. This mechanism implements a kind of *built-in model completion* that is not necessary in ordinary modal logics as they deal with only a *single* relation (viz., unlabeled arcs) that connects elements from $\mathcal{W}$ and the two modals $\Diamond$ and $\Box$ are defined in the usual dual way: $\Box\phi \equiv \neg\Diamond\neg\phi$.

5 Entailment Rules

This section addresses a restricted subset of entailment rules which will unveil new (or implicit) knowledge from graded medical statements. Recall that these kind of statements (in negation normal form) are a consequence of the application of simplification rules as depicted in Section 4.1. Thus, we assume a *preprocessing step* here that "massages" more complex statements that arise from a representation of graded (medical) statements in *natural language*. The entailments which we will present in a moment can either be *directly* implemented in a tuple-based reasoner, such as *HFC*, or in triple-based engines (e.g., Jena, OWLIM) which need to *reify* the medical statements in order to be compliant with the RDF triple model.

5.1 Modal Entailments

The entailments presented in this section deal with *plausible* inference centered around modals $\delta, \delta' \in \triangle$, some of them partly addressed in [15] in a pure OWL setting. We use the implication sign $\rightarrow$ to depict the entailment rules

$$lhs \rightarrow rhs$$

which act as *completion* (or *materialization*) rules the way as described in, e.g., [6] and [18], and used in today's semantic repositories. We sometimes even use the bi-conditional $\leftrightarrow$ to address that the LHS and the RHS are semantically equivalent, but will indicate the direction that should be used in a practical setting. As before, we define $\pi ::= p(\boldsymbol{c}) \mid \neg p(\boldsymbol{c})$.
We furthermore assume that for every modal $\delta \in \triangle$, a *complement* modal δ^{C} and a *mirror* modal δ^{M} exist (see Section 4.1).

Lift

$$(\mathsf{L}) \quad \pi \leftrightarrow \top\pi$$

This rule interprets propositional statements as special modal formulae. It might be dropped and can be seen as a pre-processing step. We have used it in the Hepatitis example above. Usage: left-to-right direction.

Generalize

$$(\mathsf{G}) \quad \delta\pi \wedge \delta \preceq \delta' \rightarrow \delta'\pi$$

This rule schema can be instantiated in various ways, using the modal hierarchy from Section 4.3; e.g., $\top\pi \rightarrow C\pi$, $C\pi \rightarrow L\pi$, or $E\pi \rightarrow U\pi$. It has been used in the Hepatitis example.

Complement

$$(\mathsf{C}) \quad \neg\delta\pi \leftrightarrow \delta^{\mathsf{C}}\pi$$

In principle, (C) is not needed in case the statement is already in negation normal form. This schema might be useful for natural language paraphrasing (explanation). Given $\triangle$, there are two possible instantiations, viz., $E\pi \leftrightarrow \neg N\pi$ and $N\pi \leftrightarrow \neg E\pi$ (note: $\mu(E) \cup \mu(N) = [0, 1]$).

Mirror

$$(\mathsf{M}) \quad \delta\neg\pi \leftrightarrow \delta^{\mathsf{M}}\pi$$

Again, (M) is in principle not needed as long as the modal proposition is in negation normal form, since we do allow for negated propositional statements $\neg p(\boldsymbol{c})$. This schema might be useful for natural language paraphrasing (explanation). For $\triangle$, there are six possible instantiations, viz., $E\pi \leftrightarrow C\neg\pi$, $C\pi \leftrightarrow E\neg\pi$, $L\pi \leftrightarrow U\neg\pi$, $U\pi \leftrightarrow L\neg\pi$, $\top\pi \leftrightarrow \bot\neg\pi$, and $\bot\pi \leftrightarrow \top\neg\pi$.

Uncertainty

$$(\mathsf{U}) \quad \delta\pi \wedge \neg\delta\pi \leftrightarrow \delta\pi \wedge \delta^{\mathsf{C}}\pi \leftrightarrow ?\pi$$

The *co-occurrence* of $\delta\pi$ and $\neg\delta\pi$ does *not* imply logical *inconsistency* (propositional case: $\pi \wedge \neg\pi$), but leads to complete *uncertainty* about the validity of π. Remember that $\mu(?) = \mu(\delta) \cup \mu(\delta^{\mathsf{C}}) = [0, 1]$ (usage: left-to-right direction):

```
      0                1
μ : |—δᶜ—|——δ——|
      π        π
```

Negation

$$\textsf{(N)} \quad \delta(\pi \wedge \neg\pi) \leftrightarrow \delta\pi \wedge \delta\neg\pi \leftrightarrow \delta\pi \wedge \delta^{\mathsf{M}}\pi \leftrightarrow \delta^{\mathsf{M}}\neg\pi \wedge \delta^{\mathsf{M}}\pi \leftrightarrow \delta^{\mathsf{M}}(\pi \wedge \neg\pi)$$

(N) shows that $\delta(\pi \wedge \neg\pi)$ can be formulated equivalently using the mirror modal:

$$\mu : \underset{\pi \wedge \neg\pi}{\overset{0}{|\!\!-\!\delta^{\mathsf{M}}\!-\!\!|}}\!\!-\!\!-\!\!-\!\!\underset{\pi \wedge \neg\pi}{\overset{1}{|\!\!-\!\delta\!-\!\!|}}$$

In general, (N) is *not* the modal counterpart of the *law of non-contradiction*, as $\pi \wedge \neg\pi$ is usually afflicted by vagueness, meaning that from $\delta(\pi \wedge \neg\pi)$, we can *not* infer that $\pi \wedge \neg\pi$ is the case for the concrete example in question (recall the intention behind the confidence intervals; see Section 3). There is one notable exception, involving the $\top$ and $\bot$ modals. This is formulated by the next entailment rule.

Error

$$\textsf{(E)} \quad \top(\pi \wedge \neg\pi) \leftrightarrow \bot(\pi \wedge \neg\pi) \rightarrow\ !(\pi \wedge \neg\pi)$$

(E) *is* the modal counterpart of the *law of non-contradiction* (recall: $\top = \bot^{\mathsf{M}}$ and $\bot = \top^{\mathsf{M}}$). For this reason and by definition, the *error* (or *failure*) modal ! from Section 3 comes into play here. The modal ! can serve as a hint to either stop a computation the first time it occurs or to continue reasoning, but to syntactically memorize the ground atoms (viz., π and $\neg\pi$) which have led to an inconsistency. Usage: left-to-right direction.

5.2 Subsumption Entailments

As before, we define two subsets of $\triangle$, called $\mathsf{1} = \{\top, C, L, N\}$ and $\mathsf{0} = \{\bot, E, U\}$, thus 1 and 0 effectively become

$$\mathsf{1} = \{\top, C, L, N, U^{\mathsf{C}}\} \qquad \mathsf{0} = \{\bot, U, E, C^{\mathsf{C}}, L^{\mathsf{C}}, N^{\mathsf{M}}\}$$

due to the use of complement modals δ^{C} and mirror modals δ^{M} for every base modal $\delta \in \triangle$ and by assuming that $E = N^{\mathsf{C}}$, $E = C^{\mathsf{M}}$, $U = L^{\mathsf{M}}$, and $\bot = \top^{\mathsf{M}}$, together with the four "opposite" cases.

Now let $\sqsubseteq$ abbreviate relation subsumption as known from description logics and realized in OWL through `rdfs:subClassOf` (class subsumption) and `rdfs:subPropertyOf` (property subsumption). Given these remarks, we define two further very practical and plausible modal entailments which can be seen as the modal extension of the entailment rules (rdfs9) (for classes) and (rdfs7) (for properties) in RDFS; see [6].

$$\textsf{(S1)} \quad \mathsf{1}p(\boldsymbol{c}) \wedge p \sqsubseteq q \rightarrow \mathsf{1}q(\boldsymbol{c}) \qquad \textsf{(S0)} \quad \mathsf{0}q(\boldsymbol{c}) \wedge p \sqsubseteq q \rightarrow \mathsf{0}p(\boldsymbol{c})$$

Note how the use of p and q switches in the antecedent and the consequent, even though $p \sqsubseteq q$ holds in both cases. Note further that propositional statements π are restricted to the positive case $p(\boldsymbol{c})$ and $q(\boldsymbol{c})$, as their negation in the antecedent will not lead to any valid entailments. Here are four *instantiations* of (S0) and (S1) (remember, $C \in \mathsf{1}$ and $E \in \mathsf{0}$):

$C\mathsf{ViralHepatitisB}(x) \wedge \mathsf{ViralHepatitisB} \sqsubseteq \mathsf{ViralHepatitis} \rightarrow C\mathsf{ViralHepatitis}(x)$
$E\mathsf{Hepatitis}(x) \wedge \mathsf{ViralHepatitis} \sqsubseteq \mathsf{Hepatitis} \rightarrow E\mathsf{ViralHepatitis}(x)$

$C\mathsf{deeplyEnclosedIn}(x, y) \wedge \mathsf{deeplyEnclosedIn} \sqsubseteq \mathsf{containedIn} \rightarrow C\mathsf{containedIn}(x, y)$
$E\mathsf{containedIn}(x, y) \wedge \mathsf{superficiallyLocatedIn} \sqsubseteq \mathsf{containedIn}$
$\rightarrow E\mathsf{superficiallyLocatedIn}(x, y)$

5.3 Extended RDFS & OWL Entailments

In this section, we will consider some of the entailment rules for RDFS [6] and a restricted subset of OWL [18]. Remember that modals only head literals π, neither TBox nor RBox axioms. Concerning the original entailment rules, we will distinguish *four principal cases* to which the extended rules belong (we will only consider the unary and binary case here as used in description logics/OWL):

1. TBox and RBox axiom schemas will not undergo a modal extension;
2. rules get extended in the antecedent;
3. rules take over the modal from the antecedent to the consequent;
4. rules aggregate several modals from the antecedent in the consequent.

We will illustrate the individual cases in the following subsections with examples by using a kind of description logic syntax. Clearly, the set of extended entailments depicted here is *not complete*.

Case-1 Rules: No Modals Entailment rule rdfs11 from [6] deals with class subsumption: $\mathsf{C} \sqsubseteq \mathsf{D} \wedge \mathsf{D} \sqsubseteq \mathsf{E} \rightarrow \mathsf{C} \sqsubseteq \mathsf{E}$. As this is a terminological axiom schema, the rule stays *constant* in the modal domain. Example:

$\mathsf{ViralHepatitisB} \sqsubseteq \mathsf{ViralHepatitis} \wedge \mathsf{ViralHepatitis} \sqsubseteq \mathsf{Hepatitis}$
$\rightarrow \mathsf{ViralHepatitisB} \sqsubseteq \mathsf{Hepatitis}$

Case-2 Rules: Modals on LHS, No or $\top$ Modals on RHS The following original rule rdfs3 from [6] imposes a range restriction on objects of binary ABox relation instances: $\forall \mathsf{P.C} \wedge \mathsf{P(x, y)} \rightarrow \mathsf{C(y)}$.
The extended version (which we call Mrdfs3) needs to address the proposition in the antecedent, but must not change the consequent (even though we always use the $\top$ modality here for typing; see Section 2):

$$(\mathsf{Mrdfs3}) \quad \forall \mathsf{P.C} \wedge \delta \mathsf{P(x, y)} \rightarrow \top \mathsf{C(y)}$$

Example: $\forall \mathsf{suffersFrom.Disease} \wedge L\mathsf{suffersFrom}(x, y) \rightarrow \top \mathsf{Disease}(y)$

Case-3 Rules: Keeping LHS Modals on RHS Inverse properties switch their arguments [18]: $\mathsf{P} \equiv \mathsf{Q}^{-} \wedge \mathsf{P(x, y)} \rightarrow \mathsf{Q(y, x)}$.
The extended version of rdfp8 simply keeps the modal operator:

$$(\mathsf{Mrdfp8}) \quad \mathsf{P} \equiv \mathsf{Q}^{-} \wedge \delta \mathsf{P(x, y)} \rightarrow \delta \mathsf{Q(y, x)}$$

Example: $\mathsf{containedIn} \equiv \mathsf{contains}^{-} \wedge C\mathsf{containedIn}(x, y) \rightarrow C\mathsf{contains}(y, x)$

Case-4 Rules: Aggregating LHS Modals on RHS Now comes the most interesting case of modalized RDFS/OWL entailment rules that offers several possibilities on a varying scale between *skeptical* and *credulous* entailments, depending on the degree of uncertainty, as expressed by the measuring function μ of the modal operator. Consider the original rule rdfp4 from [18] for transitive properties P: $\mathsf{P}^+ \sqsubseteq \mathsf{P} \wedge \mathsf{P}(\mathsf{x},\mathsf{y}) \wedge \mathsf{P}(\mathsf{y},\mathsf{z}) \rightarrow \mathsf{P}(\mathsf{x},\mathsf{z})$.

How does the modal on the RHS of the extended rule look like, depending on the two LHS modals? There are several possibilities. By operating directly on the *modal hierarchy*, we are allowed to talk about, e.g., the *least upper bound* or the *greatest lower bound* of δ and δ'. When taking the associated *confidence intervals* into account, we might even play with the low and high number of the intervals, say, by applying the *arithmetic mean* or simply by *multiplying* the corresponding numbers.

Let us first consider the general rule from which more specialized versions can be derived, simply by instantiating the combination operator $\odot$:

$$(\mathsf{Mrdfp4}) \quad \mathsf{P}^+ \sqsubseteq \mathsf{P} \wedge \delta\mathsf{P}(\mathsf{x},\mathsf{y}) \wedge \delta'\mathsf{P}(\mathsf{y},\mathsf{z}) \rightarrow (\delta \odot \delta')\mathsf{P}(\mathsf{x},\mathsf{z})$$

Here is an instantiation of Mrdfp4 dealing with the transitive relation contains from above: $C\mathsf{contains}(x,y) \wedge L\mathsf{contains}(y,z) \rightarrow (C \odot L)\mathsf{contains}(x,z)$

What is the result of $C \odot L$ here? It depends. Probably both on the application domain and the epistemic commitment one is willing to accept about the "meaning" of gradation words/modal operators. To enforce that $\odot$ is at least both *commutative* and *associative* is probably a good idea, making the sequence of modal clauses order-independent.

5.4 Custom Entailments

Custom entailments are inference rules that are not derived from universal non-modalized RDFS and OWL entailment rules (Section 5.3), but have been formulated to capture the domain knowledge of experts (e.g., physicians). Here is an example. Consider that Hepatitis B is an infectious disease

$\mathsf{ViralHepatitisB} \sqsubseteq \mathsf{InfectiousDisease} \sqsubseteq \mathsf{Disease}$

and note that there exist vaccines against it. Assume that the liver l of patient p quite hurts (modal C), but p has been definitely vaccinated (modal $\top$) against Hepatitis B before:

$C\mathsf{hasPain}(p,l) \wedge \top\mathsf{vaccinatedAgainst}(p,\mathsf{ViralHepatitisB})$

Given that p received a vaccination, the following custom rule will *not* fire (x and y below are now universally-quantified variables; z an existentially-quantified RHS-only variable):

$\top\mathsf{Patient}(x) \wedge \top\mathsf{Liver}(y) \wedge C\mathsf{hasPain}(x,y) \wedge U\mathsf{vaccinatedAgainst}(x,\mathsf{ViralHepatitisB})$
$\rightarrow N\mathsf{ViralHepatitisB}(z) \wedge N\mathsf{suffersFrom}(x,z)$

Now assume another person p' that is pretty sure (s)he was never vaccinated:

$E\mathsf{vaccinatedAgainst}(p',\mathsf{ViralHepatitisB})$

Given the above custom rule, we are allowed to infer that (h instantiation of z)

$N\mathsf{ViralHepatitisB}(h) \wedge N\mathsf{suffersFrom}(p', h)$

The subclass axiom from above thus assigns

$N\mathsf{InfectiousDisease}(h)$

so that we can query for patients for whom an infectious disease is *not unlikely*, in order to initiate appropriate methods (e.g., further medical investigations).

6 Related Approaches and Remarks

It is worth noting to state that this paper is interested in the representation of and reasoning with *uncertain assertional* knowledge, and neither in dealing with *vagueness* found in natural language (*very small*), nor in handling *defaults* and *exceptions* in *terminological* knowledge (*penguins can't fly*).

To the best of our knowledge, the modal logic presented in this paper uses for the first time modal operators for expressing the degree of (un)certainty of propositions. These modal operators are interpreted in the model theory through confidence intervals, by using a measure function μ. From a model point of view, our modal operators are related to *counting modalities* $\Diamond^{\geq k}$ [4, 1]—however, we do *not* require a *fixed* number $k \in \mathbb{N}$ of reachable successor states (*absolute* frequency), but instead *divide* the number of worlds v reached through label $\delta \in \triangle$ by the number of all reachable worlds, given current state w, yielding $0 \leq p \leq 1$. This fraction then is further constrained by requiring $p \in \mu(\delta)$ (*relative* frequency), as defined in case 5. of the satisfaction relation in Sections 4.2 and 4.3.

As [20] precisely put it: "*... what axioms and rules must be added to the propositional calculus to create a usable system of modal logic is a matter of philosophical opinion, often driven by the theorems one wishes to prove ...*". Clearly, the logic Λ is *no* exception and its design is driven by commonsense knowledge and plausible inferences, we try to capture.

Our modal logic can be regarded as an instance of the *normal* modal logic $\mathbf{K} := (N) + (K)$ when identifying the basic modal operator $\Box$ with the modal $\top$ (and *only* with $\top$) and by enforcing the *well-behaved* frame condition from Section 4.3. Given $\Box \equiv \top$, Λ then includes the *necessitation rule* (N) $p \rightarrow \top p$ and the *distribution axiom* (K) $\top(p \rightarrow q) \rightarrow (\top p \rightarrow \top q)$ where p, q being special theorems in Λ, viz., positive and negative propositional letters.

(N) can be seen as a special case of (L), the *Lift* modal entailment (left-to-right direction) from Section 5.1. (K) can be proven in Λ by choosing $\top \in \underline{1}$ in simplification rule *8b* (Section 4.1) and by instantiating (G), the *Generalize* modal entailment (Section 5.1), together with the application of the tautology $(p \rightarrow q) \Leftrightarrow (\neg p \vee q)$:

$$\frac{\dfrac{\dfrac{\dfrac{\top(p \rightarrow q) \rightarrow (\top p \rightarrow \top q)}{\top(\neg p \vee q) \rightarrow (\neg \top p \vee \top q)}}{(\top \neg p \vee \top q) \rightarrow (\neg \top p \vee \top q)}}{\top \neg p \rightarrow \neg \top p}}{\underline{\bot p \rightarrow \top^{\mathsf{C}} p}}$$

The final simplification at which we arrive is valid, since $\bot \preceq \top^{\mathsf{C}}$:

$$\mu(\bot) = [0,0] \subseteq [0,1) = \mu(\top^{\mathsf{C}})$$

Again, through (L) (right-to-left direction), Λ also incorporates the *reflexivity axiom* (T) $\top p \rightarrow p$ making Λ (at least) an instance of the system **T**. However, this investigation is in a certain sense *useless* as it does *not* address the other modals: almost always, neither (N), (K), nor (T) hold for modals from $\triangle$. Thus, we can *not* view Λ as an instance of a *poly-modal* logic.

Several approaches to representing and reasoning with uncertainty have been investigated in *Artificial Intelligence* (see [14, 5] for two comprehensive overviews). Very less so has been researched in the *Description Logic* community, and little or nothing of this research has find its way into implemented systems. [7] and [8] consider *uncertainty* in $\mathcal{ALC}$ concept hierarchies, plus concept typing of individuals (unary relations) in different ways (probability values vs. intervals; conditional probabilities in TBox vs. ABox). They do not address uncertain binary (or even n-ary) relations. [19] investigates *vagueness* in $\mathcal{ALC}$ concept descriptions to address statements, such as *the patient's temperature is high*, but also for determining membership degree (*38.5 ℃*). This is achieved through *membership manipulators* which are functions, returning a truth value between 0 and 1, thus deviating from a two-valued logic. [17] defines a *fuzzy* extension of $\mathcal{ALC}$, based on Zadeh's *fuzzy logic*. As in [19], the truth value of an assertion is replaced by a membership value from $[0, 1]$. $\mathcal{ALC}$ assertions α in [17] are made fuzzy by writing, e.g., $\langle \alpha \geq n \rangle$, thus taking a single truth value from $[0, 1]$. An even more expressive description logic, Fuzzy OWL, based on OWL DL, is investigated in [16].

Our work might be viewed as a modalized version of a restricted fragment of *Subjective Logic* [9, 10], a probabilistic logic that can be seen as an extension of Dempster-Shafer belief theory. Subjective Logic addresses subjective believes by requiring numerical values for *believe* b, *disbelieve* d, and *uncertainty* u, called (subjective) opinions. For each proposition, it is required that $b + d + u = 1$. The translation from modals δ to $\langle b, d, u \rangle$ is determined by the length of the confidence interval $\mu(\delta) = [l, h]$ and its starting/ending numbers, viz., $u := h - l$, $b := l$, and $d := 1 - h$.

Acknowledgements

The research described in this paper has been co-funded by the Horizon 2020 Framework Programme of the European Union within the project PAL (Personal Assistant for healthy Lifestyle) under Grant agreement no. 643783. The authors have profited from discussions with our colleagues Miroslav Janíček and Bernd Kiefer and would like to thank the three reviewers for their suggestions.

References

1. Areces, C., Hoffmann, G., Denis, A.: Modal logics with counting. In: Proceedings of the 17th Workshop on Logic, Language, Information and Computation (WoLLIC). pp. 98–109 (2010)

2. Baader, F.: Description logic terminology. In: Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P. (eds.) The Description Logic Handbook, pp. 495–505. Cambridge University Press, Cambridge (2003)
3. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge Tracts in Theoretical Computer Science, Cambridge University Press, Cambridge (2001)
4. Fine, K.: In so many possible worlds. Notre Dame Journal of Formal Logic 13(4), 516–520 (1972)
5. Halpern, J.Y.: Reasoning About Uncertainty. MIT Press, Cambridge, MA (2003)
6. Hayes, P.: RDF semantics. Tech. rep., W3C (2004)
7. Heinsohn, J.: ALCP – Ein hybrider Ansatz zur Modellierung von Unsicherheit in terminologischen Logiken. Ph.D. thesis, Universität des Saarlandes (Jun 1993), in German
8. Jaeger, M.: Probabilistic reasoning in terminological logics. In: Proceedings of the 4th International Conference on Principles of Knowledge Representation and Reasoning (KR). pp. 305–316 (1994)
9. Jøsang, A.: Artificial reasoning with subjective logic. In: Proceedings of the 2nd Australian Workshop on Commonsense Reasoning (1997)
10. Jøsang, A.: A logic for uncertain probabilities. International Journal of Uncertainty, Fuzzyness and Knowledge-Based Systems 9(3), 279–311 (2001)
11. Krengel, U.: Einführung in die Wahrscheinlichkeitstheorie und Statistik. Vieweg, 7th edn. (2003), in German
12. Krieger, H.U.: A temporal extension of the Hayes/ter Horst entailment rules and an alternative to W3C's n-ary relations. In: Proceedings of the 7th International Conference on Formal Ontology in Information Systems (FOIS). pp. 323–336 (2012)
13. Krieger, H.U.: An efficient implementation of equivalence relations in OWL via rule and query rewriting. In: Proceedings of the 7th IEEE International Conference on Semantic Computing (ICSC). pp. 260–263 (2013)
14. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Francisco, CA, MA (1988)
15. Schulz, S., Martínez-Costa, C., Karlsson, D., Cornet, R., Brochhausen, M., Rector, A.: An ontological analysis of reference in health record statements. In: Proceedings of the 8th International Conference on Formal Ontology in Information Systems (FOIS 2014) (2014)
16. Stoilos, G., Stamou, G.B., Tzouvaras, V., Pan, J.Z., Horrocks, I.: Fuzzy OWL: uncertainty and the semantic web. In: Proceedings of the OWLED '05 Workshop on OWL: Experiences and Directions (2005)
17. Straccia, U.: Reasoning within fuzzy description logics. Journal of Artificial Intelligence Research 14, 147–176 (2001)
18. ter Horst, H.J.: Completeness, decidability and complexity of entailment for RDF Schema and a semantic extension involving the OWL vocabulary. Journal of Web Semantics 3, 79–115 (2005)
19. Tresp, C.B., Molitor, R.: A description logic for vague knowledge. In: Proceedings of the 13th European Conference on Artificial Intelligence (ECAI). pp. 361–365 (1998)
20. Wikipedia: Modal logic — Wikipedia, The Free Encyclopedia (2015), https://en.wikipedia.org/wiki/Modal_logic, [Online; accessed 19-June-2015]

The Federated Ontology of the PAL Project
Interfacing Ontologies and Integrating Time-Dependent Data

Hans-Ulrich Krieger[1], Rifca Peters[2], Bernd Kiefer[1], Michael A. van Bekkum[3], Frank Kaptein[2], Mark A. Neerincx[3]

DFKI, the German Research Center for AI, Campus D3 2, Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany[1]

Interactive Intelligence Group, EEMCS, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands[2]

TNO, Kampweg 5, 3769 DE, Soesterberg, The Netherlands[3]

{*krieger, kiefer*}@*dfki.de*[1], {*r.m.peters, f.c.a.kaptein*}@*tudelft.nl*[2], {*michael.vanbekkum, mark.neerincx*}@*tno.nl*[3]

Keywords: knowledge representation & ontologies; Description Logics & OWL; representation of & reasoning with time-dependent data; transaction time; semantic interoperability; integration of upper & domain ontologies.

Abstract: This paper describes ongoing work carried out in the European project PAL which will support children in their diabetes self-management as well as assist health professionals and parents involved in the diabetes regimen of the child. Here, we will focus on the construction of the PAL ontology which has been assembled from several independently developed sub-ontologies and which are brought together by a set of hand-written interface axioms, expressed in OWL. We will describe in detail how the triple model of RDF has been extended towards transaction time in order to represent time-varying data. Examples of queries and rules involving temporal information will be presented as well. The approach is currently been in use in diabetes camps.

1 INTRODUCTION

In this paper, we describe ongoing work carried out in the European project PAL (Personal Assistant for a healthy Lifestyle) which will *improve child's diabetes regimen by assisting the child, health professional and parent. The* PAL *system will be composed of a social robot (NAO), its (mobile) avatar, and an extendable set of (mobile) health applications ... which all connect to a common knowledge-base and reasoning mechanism* (citation taken from the project's homepage; see `http://www.pal4u.eu`).

The **focus of this paper** lies on the construction of an integrated ontology, **PALO**, the PAL Ontology, that has been assembled from several independently-developed ontologies which are brought together by an interface specification, expressed in OWL (McGuinness and van Harmelen, 2004).[1] Within PAL, **PALO** serves as the common language which helps to interlink data, delivered from both symbolic and statistical components of the PAL system.

We will also detail how the triple data model of RDF is extended by two further arguments to incorporate temporal information in order to represent time-varying data (*transaction time*). In order to record the resulting quintuples, they can *either* be transformed into a set of semantic-preserving triples when stored in a triple repository, such as OWLIM (Kiryakov et al., 2005), by applying, e.g., W3C's N-ary relation encoding scheme (Hayes and Welty, 2006), *or* can be utilized immediately, when transferred to an n-tuple repository, such as *HFC* (Krieger, 2013). In PAL, we have opted for the latter case for various reasons. In this paper, we will also sneak a peek on the temporal entailment rules (Krieger, 2016) and queries that are built into the semantic repository hosting the data and which can be used to derive useful new information.

[1]The ontologies are publicly available for open research and to other institutions upon request; see `http://www.dfki.de/lt/onto/pal/`.

2 ONTOLOGIES

Overall, **PALO** consists of **eight** sub-ontologies, **seven** of which are truly independent and do not have knowledge of one another. **One** further ontology brings them together through the use of hand-written interface axioms, employing axiom constructors such as rdfs:subClassOf and owl:equivalentProperty, or by posing domain and range restrictions on certain underspecified properties. It is worth noting that across the ontologies, each property has been cross-classified as being either *synchronic*, i.e., property instances staying constant over time, or *diachronic*, i.e., changing over time (Krieger, 2010). This property characteristic can be used, amongst other things, to check the consistency of a temporal ABox or as a distinguishing mark in an entailment rule.

When we talk about an ontology here, we have to make a distinction between information from the TBox (terminological knowledge), RBox (general information about properties), and ABox (assertional knowledge). The TBox and RBox of the PAL domain stays constant, i.e., will *not* change over time. Only relation instances from the ABox might undergo a temporal change, e.g., the weight of a child at certain times, but *not* the birthdate.

2.1 *HFC*

HFC is a bottom-up forward chainer and semantic repository implemented in Java, comparable to popular systems such as Jena and OWLIM. HFC supports RDFS and OWL reasoning à la (Hayes, 2004) and (ter Horst, 2005), but at the same time provides an expressive language for defining custom rules, involving functional and relational variables, complex tests and actions, and the replacement of triples in favour of *tuples of arbitrary length*. The query language of *HFC* implements a subset of SPARQL, but at the same time provides powerful custom M:N aggregates, not available elsewhere. In PAL, we are using *HFC* to store universal knowledge (TBox, RBox), to query time-varying data (ABox), and to reason about temporal change. This is explicated in detail in Section 3.

2.2 Upper

PAL makes use of a minimal and stripped-down upper ontology that we have originally developed for the EU projects MUSING, MONNET, and TRENDMINER (Krieger and Declerck, 2014), showing a tri-partite division of the most general class Entity, viz., upp:Abstract, upp:Happening, and upp:Physical. Most notable for PAL is the upp:Happening representation which distinguishes between atomic upp:Situations and decomposable upp:Events, using properties such as upp:startsWith, upp:continuesWith, and upp:endsWith. This allows us to encode PDL-like processes and makes it also possible to define pre- and post-conditions. upp:Happenings are upp:basedOn upp:Entitys, upp:leadsTo other upp:Entitys, and upp:involves other upp:Agents.

2.3 DIT++

The DIT++ ontology is based on the taxonomy of dialogue acts, defined by Harry Bunt and colleagues (Bunt et al., 2012). The DIT++ taxonomy is translated into a subclass hierarchy, led by the most general class dial:DialogueAct. We have taken over the *general-purpose communicative functions* and parts of the *dimension-specific communicative functions*. The former dimension involves dialogue acts, such as dial:Request, dial:Instruct, or dial:AcceptSuggestion. The latter contains communicative acts which help to maintain a dialogue, by indicating, e.g., dial:AlloFeedback or dial:Pausing. dial:DialogueActs are equipped with several important properties, such as dial:sender and dial:addressee. A dialogue act furthermore incorporates the (shallow) semantics of a natural language utterance through property dial:frame. Property dial:follows records the temporal succession of dialogue acts, whereas dial:refersTo allows to refer back to previously introduced dialogue acts (e.g., as used in indirect speech).

2.4 Time

The time ontology basically defines the classes time:DiachronicProperty and time:SynchronicProperty, making it possible to characterize OWL properties (via rdf:type) as being able to undergo a temporal change or not (see Section 2), for instance

dom:birthdate rdf:type time:SynchronicProperty
dom:weight rdf:type time:DiachronicProperty

We have furthermore defined the property time:assign to implement the concept of an imperative, programming language variable that can change over time and whose time series needs to be recorded. Such functionality is used in PAL in the dialogue processing module (see Section 4.1).

2.5 Logic

The representation of transaction time in Section 3 needs to talk about the *truth* (= $\top$) and *falsity* ($\bot$) of statements. For this, we make use of a logic ontology which includes even more general *polarity* values, such as *don't know* (?) and *error* (!), arranged in a class subsumption hierarchy: $! \sqsubseteq \{\top, \bot\} \sqsubseteq ?$.

2.6 Domain

The domain ontology defines concepts and relation which are relevant to the PAL domain, e.g., dom:Activity (playing a game, cooking, making a diary entry), dom:Actor (child, family members, health professionals), emotional dom:Mood, or (learning) dom:Goals which progress over time (see Section 2.8). As the child (and its diabetes' history) is at the heart of the PAL project, dom:Child is consequently equipped with a large number of properties, dealing with family relationships, serious issues (hypoglycemia symptoms), hobbies, activities, or lab values. dom:LabValue bundles datatype properties relevant for the initial anamnesis and the diabetes use case, such as dom:bmi (body mass index), dom:height, or dom:bsl (blood sugar level). It is worth noting that such datatype properties usually map to custom XSD datatypes, designed for PAL (see Section 2.10).

2.7 Semantics

The shallow semantic representation in PAL is loosely build on *thematic* relations or roles (Fillmore, 1977), leading to general *verb* frames and including named arguments such as sem:agent, sem:patient, sem:theme, or sem:manner which can be found in frameworks, such as VerbNet, VerbOcean, or FrameNet (Ruppenhofer et al., 2006). These properties are defined on the very general class sem:Frame and are domain-restricted by very general classes; for instance, sem:agent and sem:patient map to the underspecified class sem:Actor. These general *docking* classes will later be interfaced with more specific classes from other sub-ontologies by means of interface axioms (Section 2.9). Even though the semantic representation is almost flat, additional roles such as sem:purpose (typed to sem:Frame) allow us to build up nested structures, say for a sentence like *OK, you will be asking* (frame: sem:AssigningRole) in a natural language quiz scenario between robot and child.

2.8 Goal

The goal ontology formalizes diabetes self-management progression and is based on the Dutch *Diabetes "weet & doe" doelen* (know & do goals) as formulated by the EADV (`http://www.eadv.nl/`). These recommendations structure knowledge and skills supposed to be obtained by the child from onset to adolescence in order to gradually increase autonomy. Thus, goals are attuned to age ranges and are divided into important topics, such as *nutrition* and *insulin*. These goals are translated into subclasses of goal:KnowledgeGoal and goal:SkillGoal, led by the superclass goal:T1DMGoal. One aim of the PAL system is to support self-management progression, by offering educational content and activities. The PAL system objectives that contribute to diabetes learning goals are defined as subclasses of goal:SupportingObjectives. Multilingual labels for Dutch, Italian, and English have been added to the goal classes as they were used in the dialogue. Properties, such as goal:hasLevel (the suggested age range) and goal:hasProgress (capturing percentage of completion) are defined on the general goal class goal:Goal. Dependencies between goals are captured via property goal:requiresAsClass which directly operates on class objects (see Section 4.2).

2.9 Pal

The PAL ontology first of all imports the previously introduced sub-ontologies, but also defines interface axioms in order to properly integrate the distributed information. This includes, e.g., restricting the domain and range of (possibly underspecified) properties or identifying (subsuming) classes and properties across ontologies. For example:

$\text{dom:Actor} \equiv \text{upp:Agent} \equiv \text{dial:Agent} \equiv \text{sem:Actor}$
$\text{dom:Goal} \equiv \text{goal:Goal} \sqsubseteq \text{upp:Event}$
$\text{goal:contributesTo} \sqsubseteq \text{upp:leadsTo}$
$\forall\, \text{dial:frame} \,.\, \text{sem:Frame}$

The *first* axiom identifies the important actor/agent classes that can be found in the various ontologies. The *second* statement makes goal:Goal (and dom:Goal) a subclass of the very general class upp:Event from the upper ontology (see Section 2.2). As a consequence, properties, such as upp:startsWith or upp:continuesWith, defined on upp:Event become available in instances of goal:Goal (*goals behave like events, occupying time*). The *third* declaration defines goal:contributesTo as a subproperty of the general property upp:leadsTo and constraints the relation signature from ⟨upp:Happening, upp:Entity⟩ to ⟨goal:SupportingObjective, goal:T1DMLearningGoal⟩. The *fourth* restriction links the underspecified dialogue act property dial:frame to shallow semantic frames (see Sections 2.7 and 4 for an example).

2.10 XSD Datatypes

Some of the datatype properties from the domain ontology utilize custom XSD types. For instance:

- *body mass index* dom:bmi, measured in xsd:kg_m2
- *blood sugar level* dom:bsl, *either* measured in xsd:mmol_L *or* xsd:mg_dL
- *diastolic blood pressure* dom:dbp, measured in xsd:mmHg

3 HANDLING TIME

This section shed some light on the representation of time-varying data in PAL and the underlying model, viz., *transaction time*. We will also look into how temporal information is utilized in queries and rules.

3.1 Metric Linear Time

In the following, we assume that the temporal measuring system is based on a *metric linear time*, so that we can compare starting/ending points, using operators, such as $<$ or $\leq$, or pick out input arguments in aggregates, using *min* or *max*. We furthermore require that time is *discrete* and represented by natural numbers. The implementation of *HFC* employs 8-byte long integers (XSD datatype `long`) to encode *milli* or even *nano* seconds w.r.t. a fixed starting point (Unix Epoch time, starting from 1 January 1970, 00:00:00). As a consequence, given a time point t, the next smallest or successor time point would then be $t + 1$.

3.2 Transaction Time

Transaction time (Snodgrass, 2000) makes use of temporal intervals in order to represent the time during which a fact is stored in the database, even though the ending time must not be known in advance. This is indicated by the wildcard **?** in the database table below which will later be *overwritten* by the concrete ending time.

We *deviate* here from the interval view by specifying both the starting time when an ABox statement is entered to the ontology, and, via a *separate* statement, the ending time when the statement is *invalidated*. For this, we exploit the polarity values $\top$ and $\bot$ from the logic ontology that we have already introduced in Section 2.5. This idea is shown below for a binary relation P. We write `P(c,d,b,e)` to denote row `<c,d,b,e>` in the database table `P` for relation P.

TIME	DATABASE VIEW	ONTOLOGY VIEW
$\vdots$	$\vdots$	$\vdots$
t_1	add: P(c,d,t_1,**?**)	add: $\top$P(c,d)@t_1
$\vdots$	$\vdots$	$\vdots$
t_2	**overwrite**: P(c,d,t_1,t_2)	——
t_2+1	——	**add**: $\bot$P(c,d)@t_2+1
$\vdots$	$\vdots$	$\vdots$

As we see from this picture, the invalidation in the ontology happens at t_2+1, whereas $[t_1, t_2]$ specifies the transaction time in the database. Clearly, the same transaction time interval for P(c,d) in the ontology can be derived from the two statements $\top$P(c,d)@t_1 and $\bot$P(c,d)@t_2+1, assuming that there does *not* exist a $\bot$P(c,d)@t, such that $t_1 < t \leq t_2$ (we can effectively query for this by employing the ValidInBetween test; see Section 3.4 for its use in a rule).

Extending ontologies by transaction time the way we proceed here gives us a means to easily encode *time series data*, i.e., allows us to record the *history* of data that changes over time, e.g., the blood sugar level of a child (see Section 2.4). The formal foundations for extending the triple model with transaction time can be found in (Krieger, 2016).

Given polarity value $\pi = \{\top, \bot\}$, the above statements

π P(c, d)@t

are written in *HFC* as *quintuples*, i.e.,

π c P d t

As we opt for a *uniform* representation, *axiomatic triples* from the TBox and RBox of an ontology need to be extended by two further arguments; for instance,

owl:sameAs rdf:type owl:TransitiveProperty

becomes quintuple[2]

true sameAs type TransitiveProperty "0"^^long

We read the above statement as *being true* ($\top$ = logic:true) *from the beginning of time* (long int 0 = "0"^^xsd:long).

[2]We sometimes omit namespaces here in order to make sure that a quintuple fits into a single paper line.

Information uploaded into *HFC* is also *backed up* by an external file. However, entailed information, obtained through successive rule applications (see Section 3.4) is not stored at all, as it can be restored through the same rules again. As a consequence, *wrongly-entered* information at time t can either be deleted directly in case no rule application has taken place since, or is deleted together with derived information from a later time $t' > t$ (like a DB rollback), followed by an application of the rules.

3.3 Queries and a Use Case

The query language of *HFC* can be seen as an extension of a subset of SPARQL towards general n-tuples. Consider the following quintuple excerpt from the ABox for *Lisa* who has undergone anamnesis at time 5544 and further lab values taken at 5577:

```
logic:true lisa rdf:type dom:Child "5544"^^xsd:long
true lisa dom:hasLabValue lv22 "5544"^^xsd:long
true lv22 dom:height "133"^^xsd:cm "5544"^^xsd:long
true lv22 dom:weight "28.2"^^xsd:kg "5544"^^xsd:long
true lv22 dom:bsl "9.0"^^xsd:mmol_L "5544"^^xsd:long
......
true lisa dom:hasLabValue lv33 "5577"^^xsd:long
true lv33 dom:weight "28.6"^^xsd:kg "5577"^^xsd:long
true lv33 dom:bsl "165.6"^^xsd:mg_dL "5577"^^xsd:long
......
```

What this example shows is that the blood sugar level dom:bsl for *Lisa* was measured using different *units* at different times (cf. Section 2.10). Given that all possible lab values will *not* be taken every time a medical examination takes place, we would nevertheless like to know the *latest* value for each individual property; for instance in our case, that *Lisa* is 133 cm tall (time: 5544), weights 28.6 kg (time: 5577), and has been measured with a blood sugar level of 165.6 mg/dL also at 5577. This information can be obtained through the following quintuple-based query which utilizes the complex aggregate GetLatestValues:

```
SELECT ?prop ?val ?t
WHERE logic:true lisa dom:hasLabValue ?labvalue ?t &
        logic:true ?labvalue ?prop ?val ?t
AGGREGATE ?measurement ?result ?time =
   GetLatestValues ?prop ?val ?t ?t
```

The meaning of SELECT and WHERE does not differ from SPARQL, except that quintuples are involved instead of triples. AGGREGATE specifies an aggregate with four input and three output arguments which sorts the result table obtained from SELECT-WHERE and headed by $\langle$?prop,?val,?t$\rangle$ according to the last fourth element ?t. It then takes the newest values $\langle$?val,?t$\rangle$ (argument 2 and 3) for each property ?prop (argument 1) and finally returns the following table:

?measurement	?result	?time
dom:height	”133”^^xsd:cm	”5544”^^xsd:long
⋮	⋮	⋮
dom:weight	”28.6”^^xsd:kg	”5577”^^xsd:long
dom:bsl	”165.6”^^xsd:mg_dL	”5577”^^xsd:long
⋮	⋮	⋮

3.4 Rules

As we have shown in (Krieger, 2016), the entailment rules for RDFS (Hayes, 2004) and OWL (ter Horst, 2005) can be extended naturally towards a treatment of time-varying data which mimics transaction time (Snodgrass, 2000). Here, we will present two such entailment rules which will derive new information for the PAL domain. The first one deal with properties and subproperties (see Section 2.9 for two such properties). The original rule rdfs7x from (ter Horst, 2005) is (we separate the *if-then* parts by writing ->):

```
?p rdfs:subPropertyOf ?q
?v ?p ?w
->
?w ?q ?w
```

This is exactly the syntax used in *HFC* for writing rules. The transaction time extension using quintuples is quite natural:

```
logic:true ?p rdfs:subPropertyOf ?q ”0”^^xsd:long
logic:true ?v ?p ?w ?t
->
logic:true ?w ?q ?w ?t
```

As we see, the underlined parts of the three clauses correspond one-to-one to the original rule and all statements are valid (first argument: logic:true). Instantiations of the first clause will be RBox axioms which will not change over time, thus we assign time 0 here, whereas changing time in the other two clauses is addressed by a coinciding logic variable ?t. The next rule does *not* have a counterpart in neither (Hayes, 2004) nor (ter Horst, 2005). It addresses a *functional property* P defined on x whose value y at time t_1 is specified differently at a *later* time t_2 by z, *without* invalidating y before:

```
true ?p rdf:type owl:FunctionalProperty ”0”^^long
true ?x ?p ?y ?t1
true ?x ?p ?z ?t2
->
error ?x ?p ?y ?t2
error ?x ?p ?z ?t2
@test
?y != ?z
?t1 < ?t2
ValidInBetween ?x ?p ?y ?t1 ?t2
```

This rule derives that P(x,y)@t_2 as well as P(x,z)@t_2 is an *inconsistent* (but *not* a *false*) statement in case P(x,y) does not get invalidated at $t < t_2$: $\bot \mathsf{P}(x,y)@t$. Whether this is the case is checked by ValidInBetween as explained before in Section 3.2. If the test succeeds, we mark the inconsistency through the use of the error modality ! (see Section 2.5) on the RHS.

4 ONTOLOGY IN USE

We have already presented an *use case* involving the ontology in Section 3.3, where a health professional is interested in obtaining the most recent lab values for a specific child. Here, we will look into two further examples.

4.1 Use Case 2: Dialogue Processing

The natural language dialogue engine in PAL utilizes sets of reactive *if-then*-like rules for the various health applications (e.g., diabetes diary, educational quizzes, sorting games). Simplified, the rules match against *general* as well as *specific dialogue situations* (= dialogue acts enriched by semantics and other information; see Sections 2.3 and 2.7) and generate continuations, describing how the dialogue proceeds. Both the matching information as well as the derived new information is grounded in time, represented by the transaction time model presented above, and stored in *HFC*. Even though the transaction time model and the ontology schema lead to a high abstraction level, *HFC* queries (Section 3.3) and rules (Section 3.4) would still be too *talkative* to be of easy use. Thus the reactive dialogue rules abstract away from things that need to be repeated over and over again (e.g., properties, such as dial:sender or dial:addressee; property chains; time). Here is an example of such a rule, a specialization of a general *answer*:

```
if (myLastDA <= @Request(Top)
  && lastDA < @Answer(Top)) {
  if (lastDA <= @Confirm(Top))
    lastDA.dialogueAct = AcceptRequest;
  else
    lastDA.dialogueAct = RejectRequest;
}
```

If the *sender*'s last *dialogue act* myLastDA is at least as specific as dial:Request (see Section 2.3) and we are given a confirmation by the addressee (stored in lastDA), the rule will assign a more specific dialogue act, viz., AcceptRequest to the field dialogueAct of variable lastDA; otherwise, RejectRequest is assigned. Even though lastDA and myLastDA look like imperative variables, they are implemented with the help of time:assign to record time series data (see Section 2.4). Furthermore, complex conditions, such as the subsumption tests above are compiled into complex SPARQL-like ASK queries.

4.2 Use Case 3: Goal Progression

The goal ontology is used to inform the child, its parents, and the healthcare professionals on the current status of self-management, but also to direct the PAL system to provide suitable content and activities. Imagine a child *Henk*, recently diagnosed with diabetes and started treatment, including self-management educational goals. *Henk* already learned that insulin intake is needed, thus goal:InsulinIntake is achieved and is given progress value 1.0. Note how the domain and goal sub-ontologies interacts (below, we omit the first argument logic:true and the transaction time argument of the quintuple in lack of space):

henk dom:hasTreatment henks_treatment
henks_treatment dom:hasGoal insulinIntake_henk
insulinIntake_henk goal:hasProgress "1.0"^^xsd:float

Henk's first selected objective is to learn to inject insulin. This requires knowledge on the location for injection and skills to prepare the insulin pen. Upon selection of goal:InsulinInjection, the progress value of this goal and its pre-conditions goal:PreparePen and goal:InsulinLocation is set to 0.0, as for related subclasses of goal:SupportingObjectives:

InsulinInjection goal:requiresAsClass PreparePen
InsulinInjection goal:requiresAsClass InsulinLocation
InsulinLocation goal:requiresAsClass AnswerI1
insulinLocation_henk goal:hasProgress "0.0"^^xsd:float
answerI1_henk goal:hasProgress "0.0"^^xsd:float
preparePen_henk goal:hasProgress "0.0"^^xsd:float

While playing a quiz, the PAL system keeps track of the scores and for each correct answer, the corresponding progress value is updated at a later time:

answerI1_henk goal:hasProgress "0.2"^^xsd:float

After correctly answering all related quiz question, the goal is achieved and all connected learning goals advance progression. Since goal:InsulinLocation has no other pre-condition, progress is updated to 1.0. As goal:InsulinInjection also specifies goal:PreparePen as a further pre-condition via property goal:requiresAsClass (see above), it is therefore progressing to 0.5 (both pre-conditions are equally important):

answerI1_henk goal:hasProgress "1.0"^^xsd:float
insulinLocation_henk hasProgress "1.0"^^xsd:float
preparePen_henk goal:hasProgress "0.0"^^xsd:float
insulinInjection_henk goal:hasProgress "0.5"^^xsd:float

ACKNOWLEDGEMENTS

The research described in this paper has been partially financed by the European project PAL (Personal Assistant for healthy Lifestyle) under Grant agreement no. 643783-RIA Horizon 2020.

Thanks to Antoine Cully and Maxime Petit (Imperial College London) for the preparation of the PAL *Data Definition*. We would like to thank the reviewers for their suggestions.

REFERENCES

Bunt, H., Alexandersson, J., Choe, J.-W., Fang, A. C., Hasida, K., Petukhova, V., Popescu-Belis, A., and Traum, D. (2012). Iso 24617-2: A semantically-based standard for dialogue annotation. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 430–437.

Fillmore, C. J. (1977). The case for case reopened. In Cole, P. and Saddock, J. M., editors, *Grammatical Relations. Syntax & Semantics 8*, pages 59–81. Academic Press.

Hayes, P. (2004). RDF semantics. Technical report, W3C.

Hayes, P. and Welty, C. (2006). Defining N-ary relations on the Semantic Web. Technical report, W3C.

Kiryakov, A., Ognyanov, D., and Manov, D. (2005). OWLIM – a pragmatic semantic repository for OWL. In *Proceedings of the International Workshop on Scalable Semantic Web Knowledge Base Systems*, pages 182–192.

Krieger, H.-U. (2010). A general methodology for equipping ontologies with time. In *Proceedings LREC 2010*.

Krieger, H.-U. (2013). An efficient implementation of equivalence relations in OWL via rule and query rewriting. In *Proceedings of the 7th IEEE International Conference on Semantic Computing (ICSC)*, pages 260–263.

Krieger, H.-U. (2016). Integrating graded knowledge and temporal change in a modal fragment of OWL. In van den Herik, J. and Filipe, J., editors, *Agents and Artificial Intelligence*, Lecture Notes in Computer Science. Springer, Berlin. Revised selected papers from the 8th International Conference, ICAART 2016.

Krieger, H.-U. and Declerck, T. (2014). TMO—the federated ontology of the TrendMiner project. In *Proceedings of the 9th edition of the Language Resources and Evaluation Conference (LREC)*.

McGuinness, D. L. and van Harmelen, F. (2004). OWL Web Ontology Language Overview. Technical report, W3C.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R., Johnson, C. R., and Scheffczyk, J. (2006). FrameNet II: Extended theory and practice. Technical report, International Computer Science Institute (ICSI), University of California, Berkley.

Snodgrass, R. T. (2000). *Developing Time-Oriented Database Applications in SQL*. Morgan Kaufmann, San Francisco, CA.

ter Horst, H. J. (2005). Combining RDF and part of OWL with rules: Semantics, decidability, complexity. In *Proceedings of the International Semantic Web Conference*, pages 668–684.

C Master Theses

Thesis Abbreviated Version

Agents Sharing Secrets: Self-Disclosure in Long-Term Child-Avatar Interaction

Franziska Burger

s4523067

Sunday 11th December, 2016

Abstract

A key challenge in developing companion agents for children is keeping them interested after novelty effects wear off. Self-Determination Theory posits that motivation is sustained if the human feels related to the agent. According to Social Penetration Theory, such a bond can be welded through the reciprocal disclosure of information about the self. As a result of these considerations, we developed a disclosure dialog module to study the self-disclosing behavior of children in response to that of a virtual agent. The module was integrated into a mobile application with avatar presence for diabetic children and subsequently used by 11 children in an exploratory field study over the course of approximately two weeks at home. It was found that the relative amount of disclosures that children made to the avatar was an indicator for the relatedness children felt towards the agent at the end of the study. Girls were significantly more likely to disclose and children preferred to reciprocate avatar disclosures of lower intimacy. No relationship was found between the intimacy level of avatar disclosures and child disclosures. Particularly the last finding contradicts prior child-peer interaction research and should therefore be further examined in confirmatory research.

1 Introduction

Social relationships often play a large motivational role in our behaviors. But we will obviously not do everything for everyone. How much we like or want to be liked by someone is an important factor. This warrants the assumption that when wanting someone to do something, effort should be invested into the bond with said someone.

According to Self Determination Theory (SDT), successful establishment of a social bond between human and agent leads to sustained motivation both to interact with the agent and to engage in activities that the agent proposes. SDT [9] argues that the basic psychological needs for *autonomy*, *competence*, and *relatedness* must be satisfied by the social environment for humans to feel motivated to attempt a task. Relatedness here refers to the feeling that one is accepted and cherished by another individual or community. It comes into play when the intrinsic motivation to engage in an activity is low. More simply put: if we like or want to be liked by someone, we feel more inclined to do what they suggest, even if we are not too fond of the activity itself.

The manner in which such a bond could be established is described by Social Penetration Theory (SPT) [1]. It proposes a directional development of interpersonal relationships whereby the involved individuals first share and explore each others personalities at a superficial level before disclosing more intimate information. Disclosing proceeds along two dimensions: breadth and depth, with *breadth* describing the number of different topics that are disclosed about and *depth* describing the personal value these topics have. Finally, an important determinant of self-disclosure is reciprocity. This describes the tendency to self-disclose as a result of being disclosed to. Reciprocal disclosures in successfully progressing relationships are usually on a similar level of intimacy.

One of the key interests in human-human self-disclosure research has been the close link between disclosure and liking. Specifically, three persistent disclosure-liking effects have been identified [8]: (a) the more someone intimately discloses to us, the more we like that person, (b) the more we like someone at the outset of the interaction, the more we will disclose, and (c) the more intimately we disclose to someone, the more we like that person.

To the best of our knowledge, no study exists that investigates these effects in child-child interaction. However, when children were asked what a friend is and what differentiates a friend from a non-friend, children older than nine indicated that friends take an interest in each others problems and care for their friend's emotional well-being. Additionally, it is argued that cooperation and the insight that each child should contribute equally to the interaction can be expected in this age group [21].

In line with this, 6th grade children's liking of another child was influenced by that child's ability to match the intimacy level of a disclosure while that of 4th graders was not [19].

Support for the disclosure-liking effect has also been found in the domains of human-robot (HRI) and child-robot (cHRI) interaction. In [17], a computer first disclosed some information about itself before asking the user an interview question. As hypothesized, interviewees shared more intimate information with the computer that told personal information about itself but only if this personal information would gradually increase in intimacy throughout the interview. However, the liking for the computer only depended on the sharing of personal information and was not influenced by the intimacy strategy. When a robot was used to elicit self-disclosures from children, those who were prompted to disclose to the robot described the robot significantly more often as a *friend* than children in the control condition [14]. In [13], a two-month study was conducted in an elementary school with a relational robot capable of identifying children and calling them by name, showing more varied behavior with time, and disclosing personal information as a function of a child's interaction time. It was found that children's desire to be friends with the robot at the end of the study was positively correlated with the interaction time.

In summary, one possibility for sustaining motivation is by leveraging relatedness. SPT provides the necessary tool for establishing relatedness: reciprocal self-disclosure with increasingly intimate content. Human-machine interaction studies further indicate that a bond between user and machine can be created through self-disclosure. Two knowledge gaps can be identified from the related literature. For one, there has been no empirical investigation of whether and how the sharing of disclosures between user and system contributes to sustaining user motivation over longer periods of time. For another thing, studies on self-disclosure reciprocity in child-child interaction have been conducted mainly in North America several decades ago (compare [7, 18, 19]). It was therefore uncertain whether insights transfer to today's children in Europe or to child-robot interaction. Furthermore, studies conducted within the framework of the ALIZ-E project[1] also showed differences between healthy and diabetic children with regard to robot interaction.

The here described research presents a first step in closing these knowledge gaps. We developed the initial prototype of a dyadic disclosure dialog module (3DM) to gain insights into how and how readily diabetic children respond to self-disclosures of an ECA and to learn about the possibilities of sustaining children's motivation in this way. A situated approach was taken by integrating the module into a mobile application for diabetic children to be used in an uncontrolled environment for a

[1] `http://www.aliz-e.org/`

period of two weeks.

The following two broad research interests guided this exploratory investigation:

1. How do children respond to a self-disclosing avatar?

2. What are the possibilities and limitations of establishing relatedness through self-disclosure and motivation through relatedness in the context of the MyPal application?

2 Development of 3DM

The first prototype of the dyadic disclosure dialog module (3DM) was developed to be integrated into the PAL-system. While it is the ultimate goal of the module to manage the sharing of personal information between agent and child in an adaptive and engaging manner, the first prototype only served the purpose of exploring the disclosure behavior of the children in interacting with an ECA. For this to be possible, it was required that there is actually content that the avatar can disclose. The first section, Section 2.1, hence details the steps taken to develop the disclosure database. This is followed by a description of how the module is integrated into the PAL-system and an explanation of the interaction flow between child and avatar as managed by the prototype in Section 2.2.

2.1 Development of the content

To design suitable disclosures for the embodied conversational agent (ECA), three preliminary steps had to be taken. First, a personality for the avatar was crafted. Second, a background story was written for the robot from which consistent disclosures at various intimacy levels could be derived. Third, a scaling method for the intimacy level of both child and avatar disclosures was developed.

2.1.1 Personality

Personality traits were selected by first choosing sensible traits for the given domain:

- *extraverted*: The ECA has to interact with many children and give presentations at camps and in the hospital. Also, it should always be very interested in its interaction partners.
- *conscientious*: Conscientiousness is very important in diabetes self-management. A conscientious ECA can provide positive examples of self-discipline and diligence for the children.
- *warm*: The ECA should function as an *opener* [16], that is, someone who evokes disclosures from the other party. To this end, it must exude trustworthiness.
- *energetic*: The ECA should encourage and motivate children to lead an active lifestyle. Additionally, it should never "not feel" like playing or chatting with a patient.

The Murphy-Meisgeier Type Indicator for Children[2] was then employed for finding a suitable type to integrate these initial traits into one coherent personality. As a result, the ECA was given the type EFJ[3]. Descriptions of this type provided insights into reasonable additional negative qualities (fear of change, inability to handle criticism, high need for praise, people-pleaser) but also additional positive qualities (determination, creativity, curiosity, cooperativeness). It can be hard for diabetic children to cope with their chronic illness psychologically. To match the child's condition, we decided to give the robot one that is not diabetes but similar in its social impact. Since NAO robots are known to overheat regularly, the pal robot was outfitted with a heat condition that regularly interferes with its lifestyle.

2.1.2 Biography

When creating the biography, the goal was to obtain a story that is both in line with the fact that robots are not human and in line with a character that children can embrace[4]. There are three main episodes to the NAO's life:

1. *Nao Nursery:* NAO robots are made in France. When they are not sold immediately, they go to the NAO nursery, which can be imagined as a big playground for robots.
   Rationale: Although the ECA is not needed somewhere in the world straight away, it is not alone. Instead, it is surrounded by many others that are its equals. It is through interactions with peers that children learn to become social beings, to compromise, to become interpersonally sensitive [21].

2. *Family:* The ECA is first acquired by a rich family. There, it experiences the novelty effect first hand. After being enjoyed as a toy for approximately one month, it is banned to the attic for two years.
   Rationale: This period was chosen to give the ECA some depth and to make children feel understood when they share negative experiences.

3. *Hospital:* The ECA was donated by the family to the local hospital. This is where it lives now together with many other care robots and the human patients of course. Here, it is well cared for.
   Rationale: Children should imagine it living in a pleasant environment where it is comfortable. They should also believe that it enjoys its daily work and especially talking to them and playing with them.

[2] `https://www.capt.org/`
[3] `https://www.kidzmet.com/blog/2015/03/08/the-extraverted-feeling-child/`
[4] `http://latd.tv/Latitude-Robots-at-School-Findings.pdf`

2.1.3 Intimacy scaling

To design agent disclosure statements at various intimacy levels and to assess the depth of children's disclosures, a rating scale for disclosure intimacy was needed. Since no adequate rating scale could be found in the literature, we created our own. This is detailed in [5].

2.1.4 Self-disclosure database

The current database consists of approximately 150 English disclosures for the avatar at all four intimacy levels. They are organized into the four categories *food*, *school*, *social*, and *sports*. These categories can be matched to those of activities that the child adds to its diabetes diary or to topics of quiz questions. In the diary environment, the child can further indicate its mood. Consequently, the disclosures have valence labels to be matched to the mood indication. In a recent study with high-school students [15], it was found that the expressivity of a robot influenced the students inclination to self-disclose. As a result, each disclosure also has an associated gesture pattern specifically for the NAO. The disclosures are stored as instances of the Disclosure class—a class in the associated ontology described in the following section. Since two of the partner hospitals of the PAL project are in the Netherlands and the study was carried out with Dutch children, all disclosures also have Dutch translations.

2.2 Development of the functionality

2.2.1 Ontology

There are three main classes in the ontology for 3DM: Disclosure, Prompt, and Closer. These correspond to the three types of statements that 3DM relies on. All disclosures have the parameters intimacy level, valence, and topic. Agent disclosures additionally have an associated gesture for the NAO robot and an associated prompt. Prompts are said by the agent to elicit a disclosure from the child. Closers are used to end the off-activity chat and return to the activity. A positive closer is said when the child chooses to disclose something, a negative closer is said otherwise. Since the module is not yet capable of comprehending a child's disclosure, closers are very general statements that make no reference to the disclosure content. The ontology is specified in RDF[5]. The relations between the classes are illustrated in Figure 1.

[5] `https://www.w3.org/`

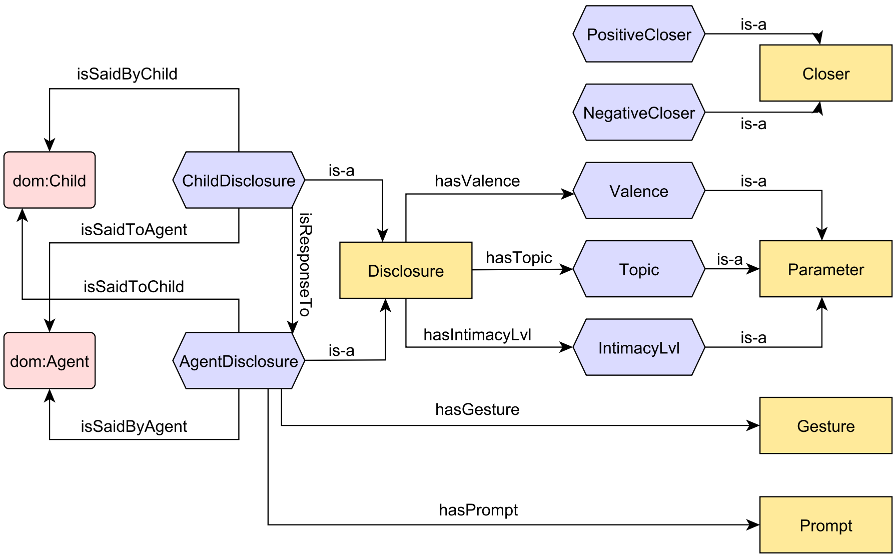


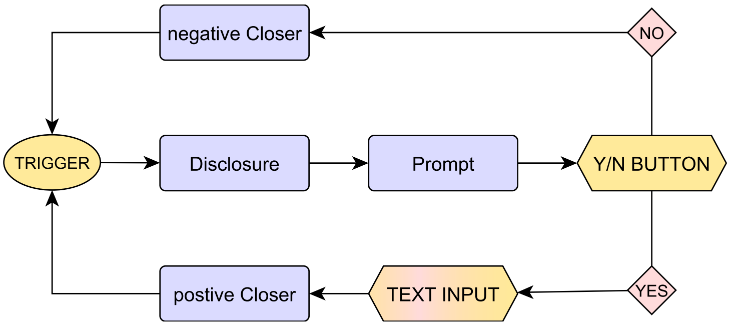
Figure 1: Ontology of the dyadic disclosure dialog module

2.2.2 Dyadic disclosure dialog module

The flow of the disclosure module follows a loop. From the perspective of the user this proceeds as illustrated in Figure 2. While inactive, 3DM waits for a trigger event from the interface. When it receives this, it selects a disclosure and sends it with a gesture to the avatar for rendering. Upon execution, it follows up with the prompt. The interface then provides a pop-up asking the child whether it would like to respond. If the child chooses not to, a negative closer command is sent to the avatar. If the child wants to respond, it can do so in a second pop-up that allows it to type some text. Once the module has received the text, it sends a positive closer command to the avatar. It then simply waits for the next trigger event. In the first prototype, the trigger event was chosen to be the opening of the diabetes diary area of the app. Both closer sentences and prompt sentences contain a placeholder for using the name of the child. It is randomly decided whether to use the name in the prompt, in the closer, or not at all.

An example dialog of the agent (**A**) with a fictional child (**C**) called Maria may look like this:

A(disclosure): "I also go to school! Together with all the other robots at the hospital. Our teachers are doctors and nurses."
A(prompt) : "Enough about me! Tell me something interesting about yourself!"
Interface : *Would you like to tell NAO something?* yes/no
C(selecting) : yes
Interface : *Please provide your response below.* text input field



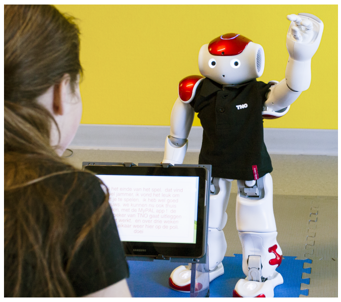


Figure 2: *Left.* Illustration of the 3DM functionality. Interface actions are hexagonal, agent actions are rectangular, and child actions are diamonds. The trigger event has a circular shape. *Right.* A diabetic child interacts with the PAL robot. Photo courtesy of Rifca Peters.

C(typing) : "I had a lot of fun at school today. We played hide and seek during the break. No one found me!"
A(p. closer) : "Thanks for sharing that with me, Maria!"

3 Method

To investigate how children behave towards the avatar, how they respond to its disclosures, how the interaction changes their feeling of relatedness, and how their motivation to use the application develops over time, a two-week, exploratory field study was conducted. The research questions are briefly repeated, before going into detail on how we strove to answer them.

3.1 Research Questions and Variables

The research questions below were of interest at the beginning of the project. However, due to unforeseen events in the course of the field study, questions $RQ2$ and $RQ3c$ had to undergo some modification. Additionally, $RQ5$ was dropped completely because the collected data was not rich enough. The necessity for and form of these changes is detailed in Section 3.5 and summarized again in Section 3.6.

After the avatar had disclosed to the child, the child was given the option to respond. For simplicity, interactions in which the child chose to respond are denoted as *active interactions* and those in which it did not as *passive interactions* from here on after.

RQ1 Do children use the application more in June than in May?
Independent Variable: evaluation time (May vs. June)
Dependent Variables: usage consistency, average amount of added content (played quiz questions and diary entries) per day and child

RQ2 How do children respond to the disclosures of the avatar?

(a) When children actively respond, can the intimacy level of the child disclosure be predicted from that of the avatar disclosure?
(b) Is there a relationship between the intimacy level of the avatar disclosure and whether children choose to respond?
(c) What (if any) role do age and gender of the children play in how intimately children respond to the avatar?

Independent Variables: disclosure intimacy of robot, age of child, gender of child
Dependent Variables: disclosure intimacy of child, response/no-response choice of child

RQ3 How does the relatedness between the child and avatar depend on:

(a) the amount of disclosures the child heard from the avatar

(b) the amount of disclosures the child made to the avatar

(c) the relatedness before the intervention

Independent Variables: number of active interactions, number of passive interactions, relatedness before the study
Dependent Variables: relatedness at the end of the study

RQ4 Is relatedness a good predictor for children's motivation to use the application?
Independent Variables: relatedness at the end of the study
Dependent Variables: consistency, amount of added content (diary entries, quiz questions)

RQ5 Is there any indication for an optimal strategy in changing the intimacy level over time? (e.g. should it gradually increase?)

3.2 Participants

Participants in the study were 11 diabetic children between the ages of 8 and 12 ($Mean_{age} = 9.91\,years$, $SD_{age} = 1.08\,years$, 6 girls). All participants had previously interacted with the MyPal application at home for 2-4 weeks in May of 2016. After this initial evaluation, children were asked whether they would like to participate again in June after some changes had been made to the avatar. Children who expressed their interest were contacted by phone in the second week of June to explain the purpose of the study and to determine a possible time to meet. This method was chosen over recruiting new children for several reasons:

1. Recruiting a sufficient number of diabetic children in the target age range with no prior PAL experience from the partner hospitals was not possible.

2. Recruiting from different sources would have taken more time than could be allotted within the time-frame of this project.

3. The prior experience allowed us to compare motivation with and without the new module within subjects. However, due to the unavailability of the module in May combined with the extensive planning that these field studies require, counterbalancing was not possible.

An important participation criterion was that children had to have been diagnosed with diabetes at least six months prior to the evaluation in May to avoid any influence of effects (psychological, lifestyle, family relations) of a recent diagnosis.

3.3 Measurements

3.3.1 Relatedness between child and avatar

It was originally intended to measure relatedness exclusively with a subset of the questionnaire from the May-evaluation. It was hoped that this would permit a comparison between how related the children felt after using the application with and without the disclosure function and hence provide a baseline measure for relatedness. The comparison could then give an indication of the added value of the module.

After administering the initial questionnaire to children, however, it became evident that it was not sensitive enough to capture different attitudes of children towards the robot. A ceiling effect was obtained on all questions regarding relatedness. As a result, *RQ3c* had to be reconsidered. Since the same ceiling effect was found on the post-questionnaire of the May evaluation, the only measure that could be linked to relatedness at the end of the May-evaluation was the usage consistency of children during the evaluation: if children were not consistent, they were probably also not feeling related to the agent and vice versa. It was therefore decided to use the May-consistency as proxy for the pre-evaluation relatedness measure if a strong correlation between June-consistency and June post-evaluation relatedness would be found.

To obtain a useful assessment of the post-evaluation relatedness, the subscales *Companionship* (how much the child enjoys spending time with the avatar), *Reliable Alliance* (how trustworthy the avatar is in terms of disclosure), and *Closeness* (how attached the child feels to the avatar and how much the child believes that the avatar reciprocates this connection) from the Friendship Qualities Scale [4] were added as additional questions to the post-questionnaire with slight modifications. The *Help* subscale was not applicable due to lack of interaction of the avatar with the physical world of the child (e.g. "If I forgot my lunch or needed a little money, my friend would loan it to me."). Similarly the *Conflict* and *Transcending Problems* subscales could not be used, because it is hardly possible for conflict to arise between child and avatar within the context of the application.

3.3.2 Intimacy of disclosures

In a post-analysis, the disclosures of the children were scaled for intimacy on the same scale as the disclosures of the avatar. This was done by two independent raters.

3.3.3 Motivation

To determine children's motivation to use the system, both indirect system usage measures and direct subjective measures were gathered. In terms of system usage, the following measures were made:

1. the number of times a child chose to respond
2. the amount of content a child added to the app while interacting (quiz questions, diary entries, and active disclosure interactions)
3. the consistency with which a child used the application. This was computed per child by dividing the number of active days (days when children interacted with the app) by the number of total possible use days.

The direct, subjective measures consisted of questions taken from the May-evaluation asking the children how much they played with the application, how much they enjoyed using it, and whether they would like to continue using it.

3.3.4 Participant traits

Age, gender, time of diabetes onset, and any comorbidities of the children could already be found in the data from the May-evaluation and did thus not need to be measured again.

3.4 Materials

3.4.1 Technological

1. Tablet Computers: A set of Lenovo tablet computers running Android was bought for the May-evaluation and further evaluations of the PAL project. Tablets were reset to factory settings after the May-evaluation and the new version of the MyPal application was installed on the tablets prior to meeting the children for the first time.
2. NAO robots: The physical robot was used for three reasons. For one, it was found throughout the study that children were not producing sufficient data with the avatar to determine how they match the intimacy level of disclosures. As a result, the real robot in the final interaction session also disclosed and asked children to reply (see Section 3.5.3 below). Also, in the ALIZ-E and PAL projects, it was found that children greatly enjoy and look forward to interactions with the robot. Thus, a final interaction with the robot served as a form of reimbursement for the children's efforts in the June-evaluation. Finally, an interaction session with the robot at the end of the study allowed the children to say goodbye to their *friend* and enabled mental closure.

3.4.2 Functional

1. MyPal Application: The app consists of three main domains—the quiz, the diabetes diary, and an overview of current and achieved diabetes-related objectives of the child. Unlike in the May-evaluation, when children in June opened the diary, the avatar started the disclosure loop provided that the child was not using the application offline.

2. Hangman Game: For the final interaction between child and robot, a hangman game was programmed with the NAO robot. This included a brief initial dialog in which the robot introduced itself. It then disclosed four times to the child, each time encouraging the child to also disclose, before moving on to the actual hangman game. Children played hangman by guessing a letter and the robot would let them know whether their guess was good or not. The word, the hangman figure, and incorrectly guessed letters were displayed on a laptop screen.

3.4.3 Questionnaires

In total three questionnaires were used in the evaluation.

1. Initial Questionnaire The initial questionnaire was administered to children in the form of a semi-structured interview. It consisted of questions concerning children's relationship to the avatar, their understanding of robots, their impression of how much they used the application in May, how much they enjoyed using the application in May, and whether they would like to continue using the application. Audio recordings of the interviews were made. The initial questionnaire was identical with the final one but excluding question 4-14.

2. Intermediate Questionnaire The intermediate questionnaire was sent to the families by e-mail approximately one week into the evaluation period. Questions regarding the new functionality and subjective impression of app usage were asked.

3. Final Questionnaire The final questionnaire was the same as the initial questionnaire plus the questions from the Friendship Qualities Scale to better assess children's feelings of relatedness.

3.5 Procedure

The procedure that was followed in this study closely resembles that of the May-evaluation. Children and their parents were contacted by phone in the second week of June to inform them of the purpose of the study, to explain the details of the procedure, and to invite them to participate again. If interested, parents were asked for their email address to receive an information letter and to then schedule an initial appointment.

3.5.1 First appointment (home).

The first appointment took place in the homes of the children. The experimenter visited each of the participating families to administer the initial questionnaire and to return the tablet computers to the children. Unlike in the May-evaluation, it was decided not to include the physical robot in the initial session. Since there was no interest in measures relating to the actual robot, it was regarded as a potentially confounding variable. Also, parents were not actively involved in this study and did not have to complete any questionnaires. After signing the consent form, children were interviewed using the initial questionnaire. Once the initial interview was complete, it was explained to the child that the app now contained a new robot with a different name (Robin). Other than that, the functionalities were the same as in the prior evaluation and they could use it without further instructions. Children were not given any guidelines as to how much they should use the application per day, because we were interested in as natural of an interaction as possible.

3.5.2 Intermediate questionnaire (remote).

After one week of using the application, the families were contacted by e-mail with a link to the intermediate questionnaire.

3.5.3 Second appointment (home).

The second appointment was similar to the first appointment. Children were again visited by the experimenter in their homes. The final questionnaire was then administered in the form of a semi-structured interview between child and experimenter. The physical robot was present in its traveling case (thus not visible) but not yet set-up during the interview. After the interview, the child was given a chance to play a hangman game with the real robot before which the robot introduced itself as Robin, telling the child that it lives in the hospital, and asking it to play a short game of story-telling to get to know each other better. In the story-telling game, the robot would make a disclosure randomly at one of the four intimacy levels and encourage the child to disclose in return. When the child was finished speaking it could say a code word to signal to the robot that it was finished. After four rounds of this interaction covering all four intimacy levels, the robot proceeded to explain the hangman game. At the end of each round, the robot would use the word that it had selected to tell another disclosure (e.g. “Hmm, the word was ‘fountain’. That reminds me of another story! One time when we were playing outside...”) and to again encourage the child to also disclose. In total, four rounds of hangman could be played but children could terminate the game after any of these rounds. Each child heard between four and eight disclosures from the physical robot. Care was taken that there was no overlap with the disclosures that the avatar had already told the child during the prior evaluation period. No sound recordings were made of this game and consequently also not of the disclosures children made during the game. Disclosures during the final interaction were recorded in the form of notes made by the experimenter.

Before the experimenter left, children were asked to return their tablets. All in all, this final session took approximately 60 minutes.

3.6 Modified research questions

As explained above, the two research questions *RQ*2 and *RQ*3*c* had to be modified. To add to the active interactions between child and ECA, the physical robot was employed as an additional “discloser” in the final interaction session. *RQ*2 was therefore changed to include the type of ECA from which the disclosure came as an influencing factor (in addition to age and gender) in the intimacy of a child’s response. From here on after, a clear distinction will therefore be made between the terms ECA, avatar, and robot in the context of disclosures: *ECA* will be used to refer to the combined disclosures coming from avatar and robot, while *avatar* will denote only those disclosures that were said within the context of the app, and *robot* will denote those at the final interaction session.

Since it was not possible to reliably assess the relatedness of children at the beginning of the June-evaluation, research question *RQ*3*c* was changed to: If there is a stong, positive relationship between usage consistency in June and relatedness at the end of the June-evaluation, are the children that feel more related to the avatar also already more consistent in their app usage in May (indicating relatedness at the beginning of the June-evaluation)?

Both these changes lead to limitations in terms of the generalizability of results. These will be discussed in Section 5. It must be emphasized that making such alterations was only accepted because of the exploratory nature of the study. In the following section, the results are presented.

4 Results

This section details the various analyses[6] that were conducted to answer the identified research questions with the data gathered in the May and June evaluations. We adopted $\alpha = 0.05$ as the significance threshold. Since it is difficult to decide whether a variable is likely to be normally distributed in the population on the basis of only 11 values (there were 11 participants in this study), it was decided to use the more conservative non-parametric test statistics whenever applicable.

4.1 RQ1: May versus June usage

To compare the app usage of children between the May and June evaluation, two different measures were used: the usage consistency (how regularly did children add content to the application?) and the average amount of added content per use day (how intensively did children use the application when they used it?). Averaging by the number of days that a specific child used the application was an important means of standardization, because the May-evaluation ran over the course of approximately 3 weeks, while the June-evaluation only had a duration of approximately 2 weeks. Furthermore, in both evaluation periods, the amount of days a specific child had access to the app varied.

Measures relating to the disclosures were not included in this comparison because they were not available in the May-evaluation. The inclusion of the quiz questions in the added content measure is debatable. Children liked the quiz very much, frequently indicating in interviews that it was their favorite part of the application. However, the game only had a limited number of questions. Since many children played through most of the questions in May already, and no new questions were added in June, it is only natural that their interest in the game was much less in June. Therefore, the better measure to compare May and June activity on is the amount of diary entries that the children made and the consistency with which they made such entries. For analyses (with and without the played quiz questions), the paired Wilcoxon signed rank test was used. The results are shown in Table 1.

4.2 RQ2: Children in dialog with the avatar

Two things were of interest when regarding how children respond to the disclosures of the ECA:

[6] All analyses and plots were made using R-Cran version 3.2.4. Heatmaps were created using MATLAB 2014a.

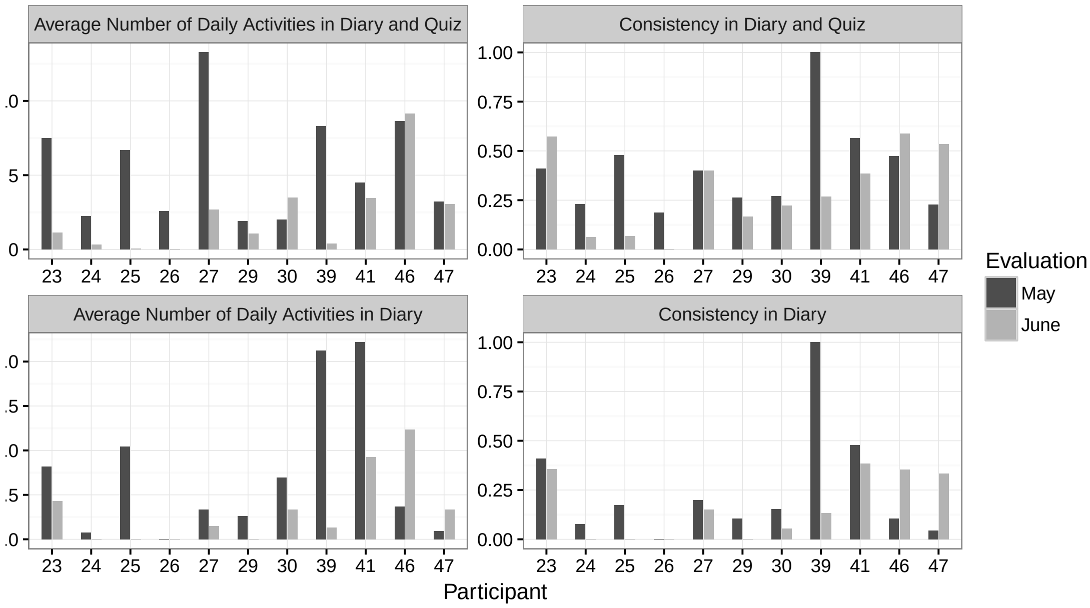


Figure 3: Visualization of activity measures in May and June for each child. The top row contains those measures pertaining to the overall usage (diary and quiz questions) while the bottom row only considered activity in the diary.

1. When children actively respond, can the intimacy level of the child disclosure be predicted from that of the ECA disclosure (taking into account age, gender, and ECA type)?

2. Is there a relationship between the intimacy level of the ECA disclosure and whether children choose to respond (taking into account age, gender, and ECA type)?

Both ECA and child disclosures were rated by two independent raters on the basis of the intimacy scale described in Section 2.1.3. Interrater agreement was assessed with a weighted Cohen's kappa. The unweighted Cohen's kappa only takes into account exact matches in ratings and is best suited when scale values are nominal and mutually exclusive. This is not the case for disclosure intimacy, which was assessed on an ordinal scale in which higher intimacy levels subsume lower intimacy levels. Hence, a weighted Cohen's kappa which squares the deviance between ratings (extent of disagreement) was employed. For the disclosures made by the ECA and the children, agreement was substantial with $\kappa = .707$, $n = 63$ and $\kappa = .697$, $n = 88$ respectively. It was therefore decided to use the ratings of one rater for further analyses. Ratings were not averaged, because this would artificially increase

Table 1: Activity comparisons between May and June evaluation based on $n = 11$ observations using a Wilcoxon signed rank test, with W and r signifying the sum of signed ranks and the effect size ($z/sqrt(2n)$) respectively.

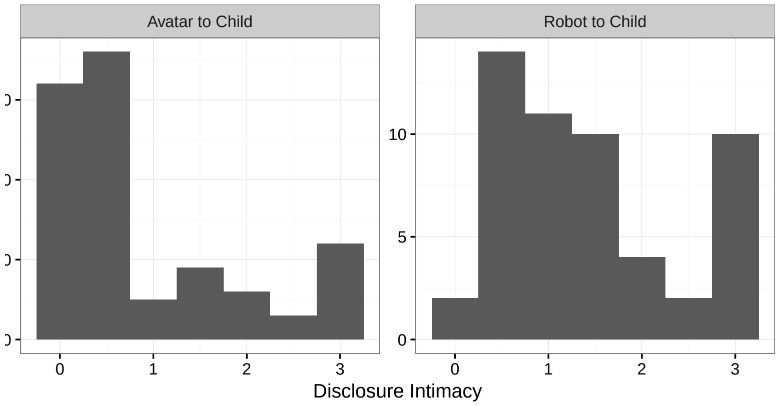
Data	**Response**	**W**	**p**	**r**
Quiz & Dairy	$\overline{Act_{day}}$	57	.032	−.65
	$Consistency$	40	.221	−.36
Dairy	$\overline{Act_{day}}$	45	.083	−.52
	$Consistency$	38	.308	−.31

the number of to-be-predicted classes and consequently decrease the number of data samples per class.

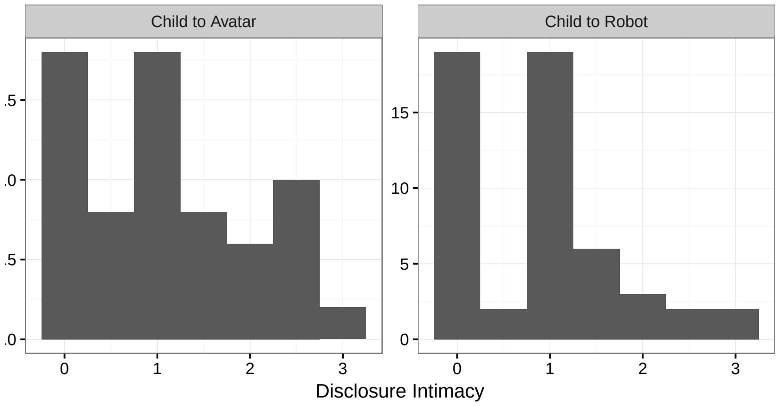
It also has to be mentioned that children did not use the application very actively resulting in sparse data. Additionally, there was a set of 'Background'-disclosures (in total 7 disclosures) that provided background information necessary for the comprehension of some other disclosures. Since they concerned just basic, factual information, they were all of very low intimacy (level 0 or 1). The avatar disclosed these before moving on to randomly select from all remaining disclosures. As a consequence of this behavior and the children's overall little usage of the application, the distribution of ECA disclosures over the various levels is not uniform. The top two rows of Figure 4 depict the various distributions of disclosure intimacy (average of both raters) from the two types of ECA and the respective response intimacies of children.

4.2.1 Child actively responds

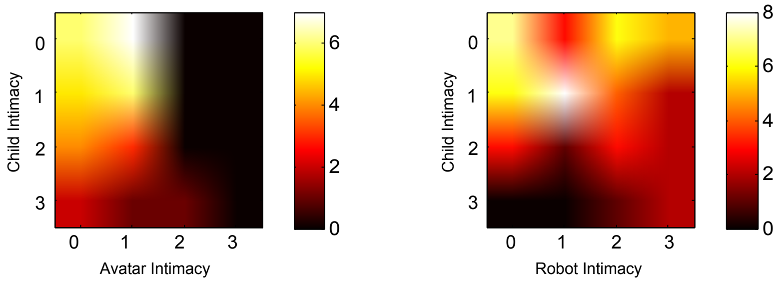
To see which effect the intimacy level of the ECA disclosure had on the intimacy of the child disclosure, linear models were fit to the data. The data is hierarchical with disclosures nested within children. As a first step, the need to use a multilevel linear model for the data was therefore determined following [10, Sec. 19.6.6.]. To this end, a model that uses the individual mean intimacy for each child ($AIC = 248.7$) was compared to the baseline model of the overall mean across children ($AIC = 247.1$) using the Akaike's Information Criterion (AIC). Since the AIC is higher for the model that allows the intercepts to vary per child, there is no variation in the data that is attributable to the random factor *child*. For the sake of a simpler model, it was therefore decided not to fit a multilevel model. Instead, a cumulative link model was chosen.



(a)



(b)



(c)

Figure 4: Figure 4a shows the distribution of disclosure intimacies separately for the avatar and the robot. This is obtained by taking the mean of both raters. Figure 4b illustrates the distributions of child intimacy in response to avatar and robot. Figure 4c shows the contingency matrix of avatar/robot disclosure intimacy and respective child disclosure intimacy as a heat map. The top left corner represents the amount of child disclosures of intimacy level 0 that were made in response to agent disclosures of level 0. Heatmap values were based on the ratings of one rater.

Several predictor variables are of interest, the most important being the intimacy level of the ECA disclosure that preceded the child disclosure. This is followed by the type of ECA (avatar or robot) that made the disclosure. The related literature indicates children's disclosure intimacy may depend on their age and gender, these variables were also included in the model. The predictors of interest were therefore: Robot.Intimacy, ECA.Type, Child.Age, and Child.Gender.

The model is given by the following equation:

$$logit(Child.Intimacy_i \leq j) = \theta_j - \beta_1(Robot.Intimacy_i) - \beta_2(ECA.Type_i) \\ -\beta_3(Child.Age_i) - \beta_4(Child.Gender_i)$$

with $i = 1, \dots, n$ and $j = 1, \dots, J$. There were $n = 88$ disclosure exchanges between the children and the robot and $J = 4$ different intimacy categories. Two assumptions are of interest for this model: multicollinearity of the predictor variables and proportional odds. Robot.Intimacy and Child.Age were not correlated ($r = .05$), the other variables are nominal. The latter assumption was assessed using the graphical method proposed in Harrell [12, p.335]. None of the predictors meet the assumption of proportional odds. To account for this, a more lenient model, allowing predictor β's to vary for each value of the outcome variable, would need to be adopted. However, this would require estimating parameters on even fewer data samples. Given the already sparse data, and the fact that there are no theoretical reasons for assuming that any of the predictor variables would affect one cumulative split of the model differently than another, it was decided to use the simpler model from the equation above. None of the independent variables played a significant role in the prediction of intimacy of child disclosure. The results are displayed in Table 2. While the model's $AIC = 227.95$ indicates a better fit to the data than the baseline model, the condition number of the Hessian is very large ($H_{cond} = 5.2e^4$). This number gives an indication of the identifiability of the model [6, p.7], with numbers larger than $1e^4$ signifying poor identifiability. This could probably be remedied by additional and more balanced data meeting the assumption of proportional odds. Prediction probabilities were not determined due to the poor fit of the model.

4.2.2 Child chooses whether to respond

Children were given the choice whether to disclose to the avatar in response to a disclosure from the avatar. It was therefore also of interest to investigate whether their choice to reciprocate depended on the intimacy level of the disclosure.

Much the same procedure as above was followed to determine the need for a multilevel linear model. Comparison of the baseline model of the mean to one allowing

Table 2: Results of fitting the cumulative link model to predict children's disclosure intimacy from the preceding ECA disclosure intimacy, the type of ECA, the age, and the gender of the child. The first five columns show the log-odds and significance tests using the Wald-statistic. The next set of three columns show the likelihood ratio if the respective predictor is dropped from the model as compared to the full model. The final three columns show the cumulative odds ratios and respective confidence intervals.

	Coefficients					Likelihood Ratio			Odds Ratio		
Predictor	*b*	*z*	*p*	*CI*		*AIC*	$\chi^2(1)$	*p*	*OR*	*CI*	
				2.5 %	97.5 %					2.5 %	97.5 %
Robot Intimacy	-.06	-.22	.829	-.60	.48	225.99	.04	.829	.94	.55	1.61
ECA Type	-.25	-.51	.610	-1.20	.70	226.21	.26	.610	.78	.30	2.01
Age	-.07	-.31	.758	-.49	.36	226.04	.10	.758	.94	.61	1.43
Gender	.41	.87	.348	-.51	1.36	226.71	.76	.383	1.51	.60	3.85

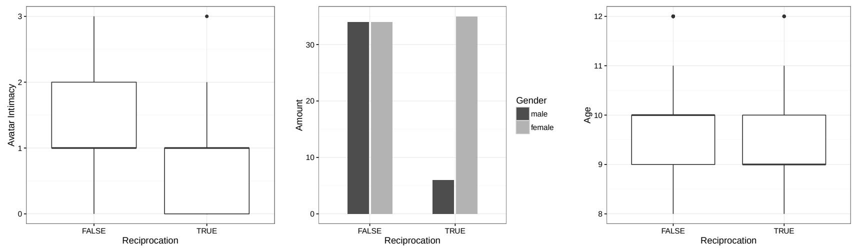
for random intercepts for each child yielded a significant improvement to fit with the latter model ($AIC_{baseline} = 155.32$, $AIC_{child} = 140.00$, $\chi^2(1) = 17.32$, $p < .0001$). Hence, a multilevel model was fit in a forced entry manner.

The multilevel model is given by the equation:

$$logit(E[Reciprocation_{i,k}]) = (\theta + \gamma_k) + \beta_1(Avatar.Intimacy_i) + \beta_2(Child.Age_k) + \beta_3(Child.Gender_k) + \beta_4(Avatar.Intimacy_i * Time_{i,k})$$

for children $k = 1, \ldots, K$ and measurements $i = 1, \ldots, n_k$ with n_k measurements per child. By adding γ_k to the intercept, the multilevel model permits different intercepts for different children. The simple logistic regression model does not include the γ_k-vector. Dropping the random effect of child ($AIC = 125.37$) and comparing to the multilevel ($AIC = 126.25$) model yielded no significant improvement ($\chi^2(1) = 2.88, p = .089$) with added complexity. As a result, the multilevel model was discarded again for the sake of a simpler model. The fit of the simple logistic regression model ($R^2 = .31$ (Nagelkerke), $AUC = .78$) was significantly better than the baseline model of the mean $\chi^2(4) = 28.10, p < .001$.

Figure 5 illustrates the effect of each predictor separately on the binary variable *Reciprocation*. The interaction term was included because the background disclosures caused disclosures of lower intimacy from the avatar to coincide with the beginning of the evaluation period. The results from fitting the model match with the visual impression. Both the intimacy level of the avatar disclosure and the gender of children significantly predict whether children choose to respond. As can be seen in Table 3,



(a) Avatar Intimacy (b) Gender (c) Age

Figure 5: The relationship between each of the predictors and the outcome variable *Reciprocation* in the logistic regression model of whether a child chooses to respond.

Table 3: Results of fitting the logistic regression model to the response choice of children within the application.

	Coefficients					**Odds Ratio**		
Predictor	b	z	p		CI	OR	CI	
				2.5 %	97.5 %		2.5 %	97.5 %
Avatar Intimacy	-.83	-1.96	.049	-1.72	-.04	.43	.18	.96
Age	.12	.51	.608	-.35	.60	1.13	.70	1.83
Gender	2.02	3.09	.002	.81	3.41	7.59	2.23	30.27
Avatar Intimacy x Time	-.00	-.15	.878	-.02	.01	.99	.98	1.01

for every unit increase in robot intimacy, the log-odds of a child disclosing decrease by .83. Furthermore, the odds of boys disclosing are 7.59 times lower than those of girls.

4.3 RQ3: Relatedness

As described in Section 1, Social Penetration Theory posits a strong link between liking and disclosure. It was hence of interest whether the disclosure activity of children was indicative of the relatedness they felt with the avatar at the end of the evaluation period.

To determine the reliability of the relatedness measure in this study, Cronbach's α was computed separately for each of the employed subscales of the Friendship Qualities Questionnaire ($\alpha_{COMP} = .73$, $\alpha_{RA} = -.41$, $\alpha_{AB} = .84$, $\alpha_{RApp} = .91$). The two items of the *Reliable Alliance* subscale were found to negatively correlate ($r = -.18$). It was thus decided to drop one of the items. For this choice, the overall Cronbach's α of all 11 items was calculated ($\alpha = .89$). Dropping the item "If there is something bothering me, I can tell my friend about it even if it is something I cannot tell to other people" increased the overall reliability of the scale ($\alpha = .90$). Active and passive disclosure counts were standardized for each child with the total number of days that it used the application.

4.3.1 Disclosure behavior and relatedness

To obtain insight into how the two different disclosure behaviors (active vs. passive) relate to the bond between child and avatar, the correlations between the variables could be determined separately. These are illustrated in Figure 6. However, these correlations do not control for the overall activity of children. The relationship between disclosure behavior and relatedness was therefore modeled using linear regression with the predictors *total number of disclosures* and *percentage of active disclosures*. The model is given by the equation:

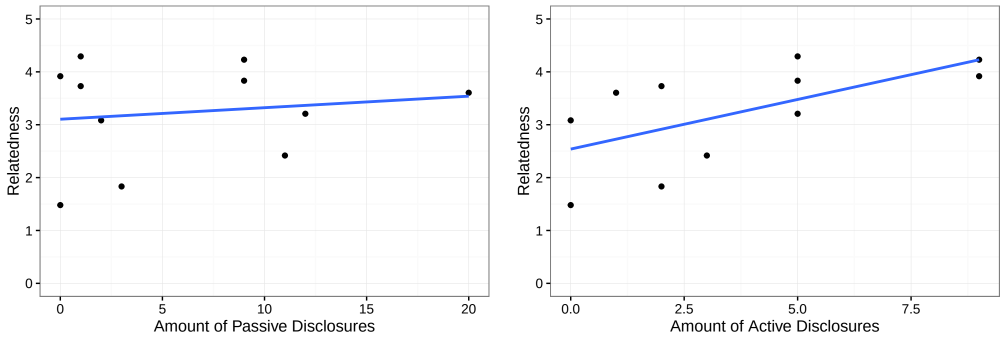
$$Relatedness = \theta + \beta_1(Disclosures) + \beta_2\left(\frac{Active.Disclosures}{Disclosures}\right)$$

The two predictors were not correlated ($\rho_S(9) = .10, p = .75$). The model (adjusted $R^2 = .45$) fits the data significantly better than the baseline model ($F(2, 8) = 5.17, p = .03$). The total amount of disclosures was not found to be a significant predictor in the model ($b_1 = 0.98, t(8) = 2.018, p = .08$). The ratio of active disclosures to total disclosures did however significantly predict relatedness ($b_2 = 1.79, t(8) = 2.690, p = .028$). This means that a unit increase in active disclosures ratio (proportionately increasing active and decreasing passive disclosures) while keeping the overall amount of disclosures constant results in a relatedness score increase of 1.79.

A problem here is causality. Since I was not able to reliably assess the relatedness of children prior to the intervention, it cannot be said whether more active disclosures lead to more relatedness or more relatedness leads to more active disclosures.

4.3.2 Relatedness and activity

Self-Determination Theory argues that relatedness plays a role in motivation. To determine whether the data of this evaluation constitute supportive evidence, the relatedness was correlated with children's overall consistency (how often they used the application) as well as their overall activity (how much they used application). Using a one-tailed Spearman's rank order correlation, a significant relationship was found between the relatedness and the consistency with which children used the application ($\rho_S(9) = .59$, $p = .03$) and the average daily activity ($\rho_S(9) = .64$,



(a) (b)

Figure 6: The relationship between the absolute amount of passive (a) and active (b) disclosures of children within the application and their relatedness as indicated on the final questionnaire.

$p = .019$). This is an indication that relatedness may positively influence motivation and even be able to uphold it over time.

To test this, a robust two-way mixed ANOVA was also carried out. For this, children were artificially split into two equally sized ($n_{related} = 6, n_{unrelated} = 5$) groups based on the overall relatedness mean. The evaluation period was divided into two halves for each child and their average daily activity (number of active contributions—diary entries, quiz questions, active disclosures—to the application per day) was calculated for each half. Thus, the relatedness constitutes the between-subjects factor and the evaluation half constitutes the within-subjects factor. Figure 7 shows the activity means of each of the $2x2 = 4$ factor level combinations. Variances were equal both across the two evaluation halves ($F(1, 20) = .12, p = .73$) as well as across the two relatedness groups ($F(1, 20) = 1.72, p = .20$). Neither main (Relatedness: $Q = .90, p = .38$; Evaluation half: $Q = 2.94, p = .17$) nor interaction effects were found ($Q = .90, p = .40$).

Since the data do not provide conclusive evidence for a link between relatedness and children's engagement with the application, children's engagement in May could not be regarded as a proxy measure for their relatedness at the outset of the June evaluation.

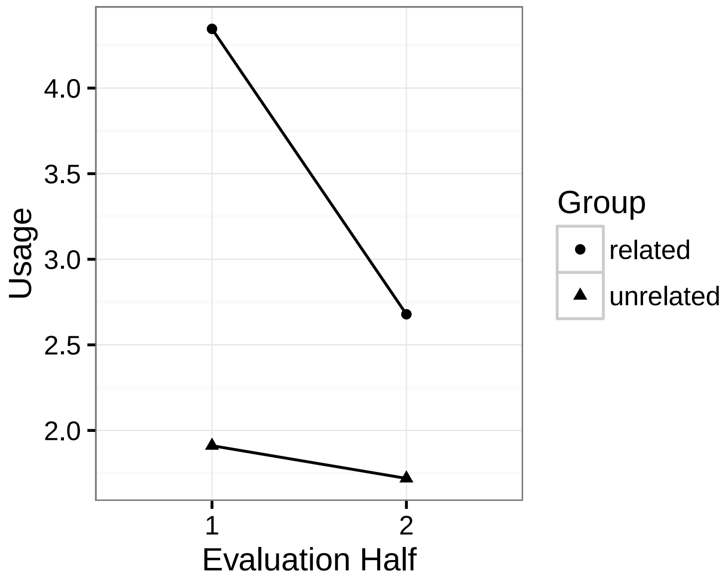


Figure 7: Average number of activities per evaluation half across children that were artificially split into the two groups related ($n = 6$) and unrelated ($n = 5$) based on their indication of Relatedness on the final questionnaire.

5 Discussion

The data analysis resulted in several interesting and partially unexpected findings. In this section, we therefore regard the results in light of the larger context of the study and its theoretical background. The nature of the research was exploratory with the goal of generating new research questions. These will be identified throughout this discussion and summarized again in Section 6.1.

5.1 Disclosure intimacy

The first matter of interest in this study was the relationship between the avatar's disclosure intimacy and whether children choose to respond as well as how they respond if they do. The former only regarded children's behavior within the diary, while the latter also included the robot.

We found avatar intimacy to be a significant predictor in whether children choose to respond with children being more responsive to disclosures of lower intimacy than disclosures of higher intimacy. This result may be limited by the confounding variable *time*. Due to the background disclosures of low intimacy that were disclosed before the robot would move to randomly select disclosures of all intimacy levels, low intimacy disclosures coincided with the beginning of the evaluation period. It is therefore possible that children disclosed more to disclosures of lower intimacy because of the novelty of the feature. Adding an interaction term of avatar intimacy and time as predictor to the logistic regression model did not improve it, indicating that time is not a moderator in the effect. Due to the small amount of data, however, it cannot be entirely excluded. If the effect is not due to the confounding variable, there are several other possible explanations. For one, children may have felt the disclosures of higher intimacy to be too much too early. It may also be that they were aware that they should match the higher intimacy but did not know anything of higher intimacy to share. The overall rather low intimacy of child disclosures that can be seen in the two heat maps in Figure 4c could be regarded as additional evidence for this. However, in the May-evaluation as well as in the focus group of the ALIZ-e project [2], parents and children stated that they would appreciate a "buddy" robot with whom children can talk about their troubles. It is therefore unlikely that children are entirely untroubled, especially when taking into consideration that they are chronically ill. Instead their troubles may not be salient enough when interacting with the app, they may not trust the avatar sufficiently despite saying so in questionnaires, or the avatar may be too limited in responsiveness. A future study could be conducted to systematically discern these possibilities.

Another significant predictor in children's decision to disclose was the gender of the child with boys making substantially fewer disclosures to the avatar than girls. Three of the five participating boys barely used the application (Participants 24, 26, and 29). Of the two boys that engaged with MyPal, both disliked the module, one because he could not get directly to the diary, the other because he did not want to talk to the avatar. For the six girls, two also showed very little usage. However, all girls expressed their liking of the module in questionnaires. Since the sample was very small, it is not clear how this generalizes to larger populations. Before drawing conclusions, the gender effect should be re-examined in a confirmatory study.

Finally, when children responded to the ECA, no pattern could be found regarding prior intimacy of the ECA's disclosure, the type of ECA, the gender, or the age of children. This contradicts prior results from child-peer disclosure behavior, in which children in the same age range as in the current study either relatively or absolutely matched the intimacy of the discloser when reciprocating [20]. From the heat maps, it appears that children are conservative in their replies, tending more towards the lower two intimacy levels regardless of the ECA's intimacy level. This result must be considered with caution, since it is based on sparse, unbalanced data. Furthermore, a problematic influence in the interactions may have been the lack of privacy given to the child when disclosing. In interactions with the physical robot, the experimenter was present and due to the spatial arrangement of some of the children's homes, it was not always possible to isolate the children from parents or siblings or ensure that no disturbances (such as family members coming home) would occur. It is also possible that children experienced similar lacks of privacy when interacting with the application or that some of the disclosures occurred in the context of children demonstrating the application to others.

All in all, the data does not paint a coherent picture with children disclosing more actively to disclosures of lower intimacy but not following any particular pattern when they do disclose. The external validity of results is not given because of the small sample size of both children and disclosures as well as the unequal distribution over different intimacy levels. Furthermore, the nature of the study led to potential influences of confounding variables. Particularly since the latter result does not match with prior findings from child-peer interaction, it is important to investigate again whether it is attributable to the replacement of the human peer with an artificial one or if other variables influenced children's true intimacy tendency.

5.2 Disclosure, relatedness, and usage

The second matter of interest was the chain of *disclosures* $\rightarrow$ *relatedness* $\rightarrow$ *motivation* that is indicated by the two human factors theories (Self-Determination Theory and Social Penetration Theory) constituting the theoretical backbone of this work. For the link between disclosure and relatedness, we found that the ratio of active disclosures to total disclosures significantly predicted the relatedness. This means that the percentage of active disclosures that children make can be regarded as an indicator for how related they feel towards the agent. A persistent finding in the related adult-adult interaction literature is that we like those more who disclose to us more [8]. This was not supported by our results, which show that it is actually the active disclosing that matters in this context. Since the initial questionnaire that we administered to children was not sensitive enough to capture their relatedness at the outset of the study, causal inferences regarding the finding cannot be made, i.e. it is unclear whether disclosing more led the children to feel more related or whether they disclosed more because they felt more related. This should be investigated again in a controlled experiment.

When regarding the link between relatedness and usage, we found no interaction effect across the two different evaluation weeks. Thus, whether children felt more or less related to the agent at the end of the evaluation did not affect their usage of the system differently in the first versus the second week. However, it must be kept in mind that the artificial split of participants into two groups means that the between-group comparisons of the robust ANOVA are based on only 5 to 6 participants. Therefore, it is more sensible to rely on visual inspection and the correlations. In so doing, we find that while more relatedness is associated with more and more consistent usage, the usage of the related group decreased substantially from the first to the second evaluation half. This is in-line with the Self-Determination Theory view on the role of relatedness in motivation, namely that relatedness is a factor in motivation, but not sufficient for it. By extension, this implies that the other two pillars of intrinsic motivation (autonomy and competence) may not be optimally met by the application. The usage curve over time from the May-evaluation supports this impression as do the claims of children in interviews and on questionnaires. While children greatly enjoy the quiz game in the application, the lack of new questions in the June-evaluation made it less attractive. The diary in the application was often stated by children as their least favorite aspect (both in May and in June). As a result, the application as a whole may not have been attractive enough for children. Several children's ideas for app improvements included the addition of new games. While this should not necessarily be taken literally, it signals children's expectation to be entertained by MyPal. While the app may compete with other apps on a very narrow market in terms of its ultimate goal (supporting diabetic children in acquiring self-management skills), the amount of applications competing for children's attention and engagement is a much larger one; one that cannot be underestimated.

Comparing the May-evaluation activity to that of the June-evaluation, children added more content in May than in June both in quiz and diary combined as well as only in the diary. The difference of the latter is not significant, but visible in Figure 3. However, children did not differ significantly in usage consistency in both evaluations. This indicates that the large amount of added content in May was mainly due to the novelty of the application. Since no control group was used, the approximately equal consistency overall between May and June evaluation cannot be attributed to the module, i.e. it is unknown whether a group of children continuing to use the application without the module would have shown a drop in consistency. Looking at the two consistency plots in Figure 3, it becomes apparent that there are large individual differences. Participant 39, for example, contributed to the application daily in May (even to the diary) but only on one-forth of the days in June. This child also indicated in the intermediate questionnaire that the module was a nuisance for him, because it prevented him from easily accessing the diary. One other child (P. 27) also remarked this. Participants 46 (youngest participant, 8) and 47 (oldest participant, 12), on the other hand, both made more diary entries more consistently in June than in May. For participant 46, this is clearly attributable to the module, because the participant pointed this out in the intermediate questionnaire and was also one of the most active disclosers. Participant 47, however, implied in the questionnaires that he did not appreciate the module much and especially did not like sharing disclosures with the avatar. It is therefore likely that additional variables that were not measured, such as more free time, contributed to his higher consistency. Thus, no clear pattern across children emerges, further supporting the need for personalization of module functionality.

In summary, it can be said that there is a link between actively disclosing and relatedness but the causal relationship needs to be further investigated. More related children did not maintain their higher initial levels of usage over time, but were using the application more than less related children.

6 Directions for further research

The nature of the study required flexibility and some adaptations had to be made to the protocol. Nonetheless, several interesting results were found. Children prefer to disclose to avatar disclosures of lower intimacy levels and girls are significantly more likely to disclose than boys. The intimacy of an ECA disclosure was a poor predictor for the intimacy of a subsequent child disclosure. Furthermore, it appears that the amount of disclosures that children make towards the avatar is an indicator of how related they feel towards it. No support could be found that children feeling more related to the avatar maintain their initially high usage over time. All findings should be addressed again in confirmatory studies.

6.1 New research questions

An important goal of this research was the generation of new research questions. These questions can be derived from both the significant and the insignificant results of this study:

nRQ1 What is the causal link between active disclosing and relatedness in the context of long-term child-avatar interaction?

nRQ2 Are children less likely to respond to more intimate avatar disclosures? If so, why?

nRQ3 Is there a general or child-dependent strategy that the ECA should follow in terms of intimacy development over time to obtain more active disclosures from children?

nRQ4 Do boys disclose less to an avatar than girls? If so, why?

nRQ5 Do children also not match the intimacy level of an ECA when they are given complete privacy?

nRQ6 Is there a difference in how children match disclosure intimacy depending on whether a physical ECA, virtual ECA, or another child is disclosing first?

nRQ7 Do children feel more related to a more responsive avatar in the context of long-term interaction?

nRQ8 Is there a difference between diabetic and healthy children in their disclosure behavior towards an ECA?

These research questions should be addressed in confirmatory studies with larger populations of children. Since the artificial intelligence field in dialog is currently still too limited to investigate the possibilities of responsive ECAs using an autonomous avatar, Wizard-of-Oz techniques could be resorted to. Regarding $nRQ3$, an interesting approach may be to reward the agent for every active disclosure it receives from the child and to have it learn the best intimacy strategy. Final strategies could then be compared across children. This, however, would require more intense application usage from the children. The module in itself is flexible and could easily be integrated into another software as well. In its current state, however, it is still too limited to provide engaging dialog interactions for children. Hence, a second prototype should be developed.

6.2 Prototype iteration

Several points of improvement for the module became evident during the study. For one, as already identified in Section 5, not all children appreciated the placement of the module within the app. This is something that seems to clearly be a personal preference and thus should be personalized. This could be done by providing a quick dialog exit option and learning the child's placement preferences (possible options include: in the quiz, in the diary itself, or after an initial greeting when opening the application).

The application was also very limited in its dialog capabilities and from the responses of children it is clear that they figured this out soon (e.g. children attempted to ask the avatar questions several times). In a recent study, participants asked to disclose a negative event to a robot rated it as more sociable, displayed more attachment manifestations, and expressed greater interest in having it as a companion when the robot was responsive to the disclosure than when it was not [3]. The authors consequently argue that responsiveness is essential in emotional bonding. Furthermore, Gottman [11] provides an example of purely disclosure-based dialog, arguing for its unnaturalness. When we interact with others, we typically do not only self-disclose. Instead, we ask questions or comment on what the discloser has said.

In a similar vein, 8 of 11 children had the impression that the avatar knew them better as a consequence of their disclosure. It would be nice for future iterations of the module if the avatar could also show this. To this end, the PAL user model should be augmented with information filtered from the dialog and means should be found to incorporate knowledge from the user model again into the dialog.

All in all, this can be summarized as a need for more intelligent behavior of the module. Ultimately and ideally, very intimate disclosures of 3DM could be triggered when it *senses* that something is the matter with the child (for example, by parsing the diary entries of the child or employing emotion recognition techniques), while disclosures of lower intimacy could be triggered by content, i.e. the trigger event and selected disclosure should be context dependent.

7 Conclusion

Due to the lack of recent research in the areas of child-peer and child-robot bonding, we conducted an exploratory field study using the first prototype of the dyadic disclosure dialog module. The purpose of the study was two-fold: on the one hand, we wanted to learn about diabetic children's behavior towards a self-disclosing virtual agent. On the other hand, we were interested in possibilities and limitations of creating a bond between child and agent to increase children's motivation in using the application. More related children both disclosed more actively and used the application more than less related children. Future research will need to investigate whether there is truly a difference between ECA and human as conversational partner for children in terms of the reciprocation of intimacy. We thus conclude that the current project presents only a starting point, but a promising one at that.

References

[1] I. Altman and D. Taylor. Social penetration theory. *New York: Holt, Rinehart &\ Mnston*, 1973.

[2] I. Baroni, M. Nalin, P. Baxter, C. Pozzi, E. Oleari, A. Sanna, and T. Belpaeme. What a robotic companion could do for a diabetic child. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 936–941. IEEE, 2014.

[3] G. E. Birnbaum, M. Mizrahi, G. Hoffman, H. T. Reis, E. J. Finkel, and O. Sass. What robots can teach us about intimacy: The reassuring effects of robot responsiveness to human disclosure. *Computers in Human Behavior*, 63:416–423, 2016.

[4] W. M. Bukowski, B. Hoza, and M. Boivin. Measuring friendship quality during pre-and early adolescence: The development and psychometric properties of the friendship qualities scale. *Journal of social and Personal Relationships*, 11(3): 471–484, 1994.

[5] F. Burger, J. Broekens, and M. A. Neerincx. A disclosure intimacy rating scale for child-agent interaction. In *International Conference on Intelligent Virtual Agents*, pages 392–396. Springer, 2016.

[6] R. H. B. Christensen. ordinal—regression models for ordinal data, 2015. R package version 2015.6-28. http://www.cran.r-project.org/package=ordinal/.

[7] N. B. Cohn and D. S. Strassberg. Self-disclosure reciprocity among preadolescents. *Personality and Social Psychology Bulletin*, 9(1):97–102, 1983.

[8] N. L. Collins and L. C. Miller. Self-disclosure and liking: a meta-analytic review. *Psychological bulletin*, 116(3):457, 1994.

[9] E. L. Deci and R. Ryan. Overview of self-determination theory: An organismic dialectical perspective. *Handbook of self-determination research*, pages 3–33, 2002.

[10] A. Field, J. Miles, and Z. Field. *Discovering statistics using R*. Sage publications, 2012.

[11] J. Gottman. How children make friends. *Monographs of the Society for Research in Child Development*, 48(3), 1983.

[12] F. E. Harrell. Regression modeling strategies. 2001. *New York: Spring-Verlag*, 2001.

[13] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro. A two-month field trial in an elementary school for long-term human–robot interaction. *Robotics, IEEE Transactions on*, 23(5):962–971, 2007.

[14] I. Kruijff-Korbayová, E. Oleari, A. Bagherzadhalimi, F. Sacchitelli, B. Kiefer, S. Racioppa, C. Pozzi, and A. Sanna. Young users' perception of a social robot displaying familiarity and eliciting disclosure. In *Social Robotics*, pages 380–389. Springer, 2015.

[15] N. Martelaro, V. C. Nneji, W. Ju, and P. Hinds. Tell me more: Designing hri to encourage more trust, disclosure, and companionship. In *The Eleventh ACM/IEEE International Conference on Human Robot Interation*, pages 181–188. IEEE Press, 2016.

[16] L. C. Miller, J. H. Berg, and R. L. Archer. Openers: Individuals who elicit intimate self-disclosure. *Journal of Personality and Social Psychology*, 44(6): 1234, 1983.

[17] Y. Moon. Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of consumer research*, 26(4):323–339, 2000.

[18] K. J. Rotenberg and N. Chase. Development of the reciprocity of self-disclosure. *The Journal of genetic psychology*, 153(1):75–86, 1992.

[19] K. J. Rotenberg and L. Mann. The development of the norm of the reciprocity of self-disclosure and its function in children's attraction to peers. *Child development*, pages 1349–1357, 1986.

[20] K. J. Rotenberg and D. Sliz. Children's restrictive disclosure to friends. *Merrill-Palmer Quarterly (1982-)*, pages 203–215, 1988.

[21] J. Youniss. *Parents and peers in social development: A Sullivan-Piaget perspective*. University of Chicago Press, 1980.

Multimodal Fission with a Focus on Collaborative Human-Robot Interaction

Magdalena Kaiser
s8makais@stud.uni-saarland.de

ABSTRACT

The aim of my master thesis is to create an extendable, domain-independent framework for **M**ulti**m**odal **F**ission (abr. MMF) with a special focus on humanoid robots as execution environment. The term Fission describes the process of selecting a combination of channels which should be used to output a specific information to the user. This paper will give an overview of a first version of the MMF framework. How to categorize modalities and how to realize the modality selection process is investigated, as well as how to enable multimodal references. Since the focus of this thesis is on Human-Robot Interaction, modalities like Speech, Pointing and Gaze are implemented. However, other modalities can be added easily. The framework receives a simple predicate as input and outputs a multimodal representation of the information which can be presented by a robot.

General Terms/Keywords

Human-Robot Interaction, Framework, Fission, Output Multimodality, Multimodal Presentation Planning

MOTIVATION

In contrast to Human-Computer Interaction, communication among people is diverse and happens in multiple ways, for instance by using speech, gestures, gaze or facial expressions. However, over the last two decades, *Multimodal User Interfaces* were developed [1], which aim to enable the usage of computers in a more natural way or more specifically, in a way that we know from communication with other humans. *Multimodal* means that several modalities are used, where a *modality* is a way to convey information. Examples for modalities are Speech, an Image or a Ringing Sound. Different modalities can be chosen to present the information to users with different language skills, age or disabilities and in varying environments. It is up to the user to choose the modalities he or she wants to use for the input, whereas in most cases, the system itself must decide which combination of channels to use to output a specific information to the user. The latter task is called *Fission*. For enabling human-like conversations with robots, *Multimodal Dialog Systems* are needed. A Multimodal Dialog System typically consists of a a *Fusion Module*, a *Dialog Manager* and a *Fission Module*. A simplified representation of such a system can be found in Figure 1. First, the Fusion Module processes the user input. Then the Dialog Manager, which is the core component of the Dialog System, extracts the meaning of the user input and tries to find a corresponding response. For this task, it usually retrieves information from a knowledge base and several other

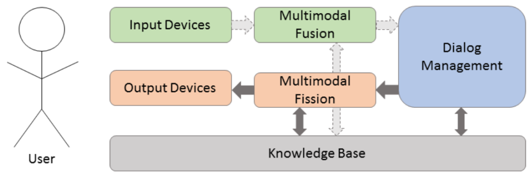


Figure 1. Simplified Representation of a Multimodal Dialog System

components, like a discourse model or a user model. In the case of proactive systems, an input is not necessary for the system to perform a communicative act. Finally, the Fission Module presents the output. It receives an abstract information from the dialog manager as input and outputs a plan, prescribing which concrete information should be presented using which modalities at which point in time during the execution. According to Costa and Duarte [1], not much research has been conducted about Fission, because most applications use only few different output modalities and thus simple and direct output mechanism are often sufficient. Moreover, the structure of most systems seems to be very dependent on their used domain.

Therefore, a **M**ulti**m**odal **F**ission framework (abr. MMF framework) will be developed in this master thesis. The MMF framework will receive a simple predicate as input and outputs a multimodal plan which can be executed by a robot to communicate with humans. Further details about the framework will be provided in the section *Description*. The following research questions will be answered in this thesis:

- How to create an extendable, domain-independent Multimodal Fission framework?
- How to categorize the modalities used in this framework?
- How to realize Multimodal References and how to resolve references via speech?
- How to realize Modality Selection and Device Selection?

Preliminary answers to these questions will be given in the section *Discussion*.

RELATED WORK

Research on multimodal output representation has already started in the early 90's. For example, a system called WIP [8] was developed at DFKI which generated illustrated text customized for the intended audience and situation.

In 2002, Foster summarized the state of the art in Fission for the COMIC project, in which a Multimodal Dialog System was developed. Additionally, she compared several systems that produce multimodal output [3]. Foster divided the tasks in Fission into three categories, namely *content selection and structuring*, *modality selection* and *output coordination*.
She defined *content selection and structuring* as the task of choosing which content should be included in the presentation and of arranging its overall structure. In general, it is the task of the Dialog Manager to determine which content should be presented to the user. However, some systems consider this task in their Fission Module. Most often, such systems use a plan-based approach to achieve this task.
However, the MMF framework receives a predicate as an input from a Dialog Manager, which directly states which content should be expressed. Therefore, *content selection and structuring* is not considered in the MMF framework (see subsection *Input and Output*).
Modality selection describes the task of choosing among all available modalities those modalities which are most suitable for realizing the particular output. To perform modality selection, knowledge about the characteristics of the available output modalities and the information to be presented can be taken into account. Furthermore, user characteristics, the performed task, communicative goals of the presenter and any resource limitations might be considered as well. Foster mentioned three possible approaches to solve this task: plan-based, rule-based and the usage of competing and cooperative agents. The modality selection process is a core component in the MMF framework (see subsection *Modality and Device Selection*).
According to Foster, the third task of a Fission Module is *output coordination*, in which the different output channels need to be coordinated so that the output forms a coherent presentation. She defines three sub-tasks: *physical layouting* (necessary if several modalities use a screen for representing information), *temporal coordination* (necessary if several dynamic modalities like speech or animations are used) and the representation of *referring expressions*. Referring expressions can be further divided into *Multimodal References* (making references using multiple modalities) and *Cross-modal References* (referring to other parts of the presentation).
Currently, no modality which requires a screen as output device is implemented in the MMF framework. Therefore, physical layouting is not considered for now. However, temporal coordination as well as Multimodal References are covered (see subsections *Categorization of Modalities* and *Reference Resolution via Speech*).

During the SmartKom project [7] in 2003, a Multimodal Dialog System was developed in which speech, gaze and facial expressions of an animated character are used. Moreover, the term *Symmetric Multimodality* was introduced to describe systems for which all modalities used for the input are also available for the output, and vice versa. In SmartKom, a plan-based approach is performed in the Fission Module. A presentation planner is used, which receives a modality-free representation of the system's intended communicative act as input. By using presentation parameters that encode user preferences, the presentation planning process can be adapted to various application scenarios.

Foster and White described their plan-based Fission approach for the COMIC project in 2005 [4]. The COMIC Dialog System is used to realize an intelligent bathroom designer. The output is presented via a GUI and a virtual talking head, which is able to speak, do facial expressions and make deictic references by gaze shifts. In order to find an output which is adequate for the current situation and the current user, the dialog history and the user's preferences, which are contained in the user model, are taken into account.
Realizing an individual output per user and taking the dialog history into account is also an aim of the MMF framework (see subsection *Modality and Device Selection*).

In 2006, Rousseau et al. presented their ELOQUENCE platform [6] and proposed a conceptual model for multimodal presentation called WWHT, which is based on the following four concepts: **W**hat information to present, **W**hich modalities to choose to present the information, **H**ow to present this information using these modalities and **T**hen - how to handle the evolution of the resulting presentation. Apart from the last, these concepts can be mapped to the concepts defined by Foster, which have already been mentioned. The last concept represents the problem how to react if the interaction context changes during the presentation. This is mainly important for persistent presentations.
Furthermore, Rousseau et al. distinguish three types of interaction components: *mode*, *modality* and *medium*. Modes correspond to human sensory systems (visual, auditory, tactile, etc.), a modality is defined by the information structure as it is perceived by the user (text, image, vibration, etc.) and a medium is a device used for the output (screen, speaker, vibrator, etc.).
In the MMF framework, the concepts of modality and medium are used. However, the term *device* is used instead of *medium*.

In the scope of the European project GUIDE, Costa and Duarte developed a Multimodal User Interface for elderly and differently impaired users in 2011 [1]. This work also focuses on adapting the choice of output representations to a corresponding user.

DESCRIPTION

The following section will provide an overview of the Multimodal Fission framework. Figure 2 gives a high-level representation of its structure.

Input and Output

The user of the framework has to decide which modalities and which of the available output devices should be offered. The list of available devices limits the list of possible modalities. Since the framework focuses on Human-Robot Interaction, this information becomes part of the *robot model*. Additionally, the robot model contains information about the robots in the environment. Furthermore, the framework requires a representation of the environment called *world model* as input. The world model contains the aforementioned robot model, alongside a *user model*, a *context model* and so called *visibility annotations*. The user model and the context model contain

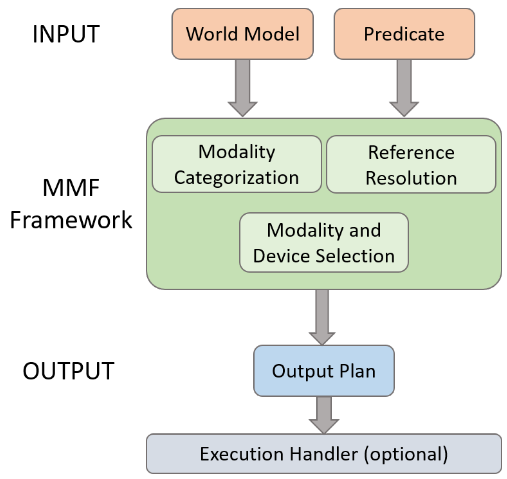


Figure 2. Simplified Representation of the MMF Framework

information about the users respectively the concrete objects in the current environment. The visibility annotations state how noticeable the objects' properties are for the user. These annotations will be explained later in more detail. The world model represents all physically present objects in the environment, including users and robots. However, it does not contain abstract concepts like a task description. Such concepts are usually defined in the knowledge base and are relevant for the Dialog Manager but not for the Fission Module. All objects, users and robots defined in the world model require a unique id, called *worldobjectid*, which is internally used to identify the object, and a *worldobjecttype*, which assigns each object a category. Additionally, position coordinates are required if certain modalities, like a Pointing Modality, are used. Furthermore, the properties of the objects are represented as key-value pairs, which can be nested if desired. For example, for a blue, small vase, the information saved in the model could look as follows: $[worldobjectid : Vase1]$, $[worldobjecttype : Vase]$, $[position : [xPos : 3.0], [yPos : 5.0], [zPos : 2.0]]$, $[color : blue]$, $[size : small]$. If the "world" described in the world model does not change during the usage of the framework, the world model must only be defined once, before the first usage. Currently, the models can be provided in form of an ontology using the Web Ontology Language (OWL)[1] or as database entries in a MongoDB[2] database. The information contained in the world model is internally stored as JSONObjects.

In addition to the world model, the MMF framework receives a *predicate* as input, encoding the information that should be presented to the user. The predicate is the input, which needs to be provided by a potential Dialog Manager. Such a predicate consists of a verb and the parts of a phrase which depend on that verb, called its *arguments*. The semantics of the predicate is based on the semantics of the corresponding predicate in the artificial language Lojban[3] (pronounced ['loʒban]). Lojban is based on predicate logic and has an unambiguous grammar. It defines for each predicate the semantics of its arguments. The advantage of using Lojban based predicates is that, if new predicates are to be inserted, Lojban's predicate structure can be used as a reference. Therefore, all predicates are built in the same way. The MMF framework currently supports 30 predicates. As an example, consider the predicate *nitcu*(x_1, x_2, x_3), which means "to need" and where the semantics of its arguments is the following: "x_1 needs necessity x_2 for purpose/action/stage of process x_3". A possible input for the MMF framework would be *nitcu(User1, Scissors2, craft task)*.

The system's output is a sequence of triples that consist of the selected modality, the selected device and the output element. For the example's input predicate mentioned above, assume there are three modalities, namely a Speech, a Pointing and a Gaze Modality, available. Furthermore, four devices, namely a Speech Device, a Left Pointing Device, a Right Pointing Device and a Gaze Device, are offered to execute the output. A possible output is presented in Figure 3. Only the Left Pointing Device has been chosen from among the Pointing Devices for this output. Some parts of the output are executed in parallel. For example, the Speech Device outputs the string "User with name Magdalena" at the same time while the Gaze Device looks at the concrete position, presumably the location of the mentioned user. Therefore, the output can be seen as a sequence of parallel actions. Additionally, the MMF framework offers an *Execution Handler*. This handler can be connected to the available devices to execute the output.

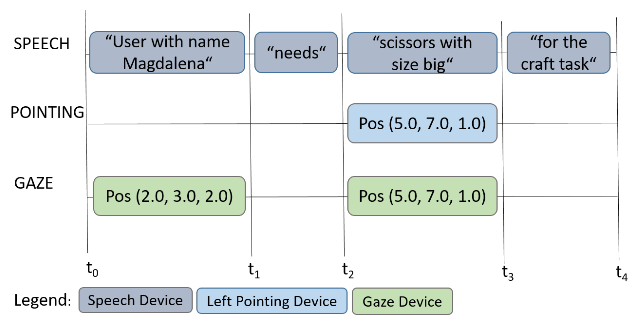


Figure 3. A possible Multimodal Output

Categorization of Modalities

For now, three modalities are available in the MMF framework: Speech, Gaze and Pointing. A list of devices is assigned to each modality. For example, the Pointing Modality receives, in the case of a humanoid robot, the robot's arms as concrete devices. In the framework, modalities are categorized into two different types: *Structure-Forming* and *Object-Referencing* Modalities. Structure-Forming Modalities determine the structure of the output and thus the order of the output elements, whereas Object-Referencing Modalities denote modalities which can make references to objects in the environment. Pointing and Gaze Modalities are examples for Object-Referencing Modalities. Objects are referenced by pointing respectively looking at the object's absolute position, which can be retrieved from the world model. The Speech Modality belongs to both categories. It has the task to create

[1]OWL: https://www.w3.org/TR/owl-ref/
[2]MongoDB: https://www.mongodb.com/
[3]Lojban-English Dictionary: http://tiki.lojban.org/tiki/tiki-download_wiki_attachment.php?attId=711

a sentence out of the predicate input, but can also reference objects in the environment.

Reference Resolution via Speech

In order to reference objects, the Speech Modality has to find an appropriate linguistic description for the object. In the above example, the desired speech output should not contain the respective *worldobjectid*, like "User1" or "Scissors2". These are abstract representations used to identify the referenced objects internally and are not sufficient for the user to recognize which object is referenced. As a solution, the object can be described by using a subset of its properties defined in the world model. Each world object has a *worldobjecttype* which could be used for the speech output. However, the *worldobjecttype* might not be a unique identifier, since there could be several objects of the same type in the environment. Nevertheless, it is assumed that objects of the same type differ in some of their properties, otherwise they would be indistinguishable and a differentiation is neither possible nor needed. For the extraction of a linguistic description for a world object, the *Unique Identifier Algorithm* was designed in the context of this thesis. The algorithm receives the object for which a unique representation should be found (in the following referred to as the *queried object*) and all other objects from the same type, including their properties, as input. As an additional input, the visibility annotations of the world model, which have been mentioned earlier, are used. These annotations need to be defined by the user of the framework in advance. He or she can assign a so called *visibility value* between 0.0 and 1.0 to each property of the world model. This value describes how noticeable the object is for the user if the corresponding property is used to identify it. For example, color is usually very noticeable (visibility: 1.0), whereas the object's absolute position in the internal coordinate system of the world is not a helpful information for the user and therefore receives visibility 0.0. In Figure 4, an example of the

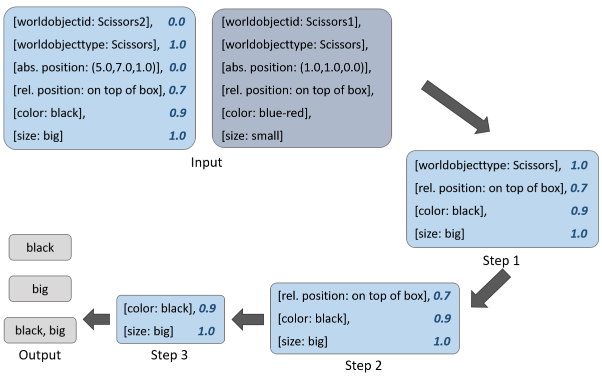


Figure 4. Example Usage of the Unique Identifier Algorithm

algorithm's usage is provided. In this case, the queried object is "Scissors2". The second input is the object "Scissors1" and the third are the visibility annotations for the scissors' properties. Only properties which have a visibility value higher than a certain threshold are considered by the algorithm. In the example, this threshold is 0.5. Properties which have a visibility value below the threshold are removed in Step 1 of the example. The algorithm compares the remaining properties of the queried object to the properties of all other objects successively. Then, it extracts those properties which differentiate the compared objects. Since there are several scissors in the example, the *worldobjecttype* cannot be used for differentiation and is removed (Step 2). In Step 3 the relative position property is removed, because both scissors have the same value for this property. The algorithm outputs a list of all possible combinations of properties which can be used to uniquely identify the queried object. The final combination of properties used for the output can be chosen from this list by using one of several criteria. One such criterion could be to use the combination with the fewest properties (black or big would be chosen in the example of Figure 4). Another one might use the combination which maximizes the *visibility value* (combination of black and big in Figure 4).

Modality and Device Selection

After the predicate is divided into its components, the modalities evaluate how well they can present the particular component by assigning a number between 0.0 and 1.0, where 0.0 means *not presentable at all* and 1.0 means *perfectly presentable*. If the number is below a threshold, the modality will not be considered for this element in the planning process. The constraint satisfaction solver OptaPlanner [2] is used for planning which modalities and devices to use for each concrete output element. OptaPlanner uses scoring calculators for planning. The user of the framework can provide their own scoring calculators based on different planning criteria. For now, selecting the output device works as follows: The Speech Device closest to the user is used for the speech part of the output and for the Object-Referencing Modalities the device located closest to the referred object is used. However, different and more sophisticated criteria can be added.
In the following, criteria for modality selection will be discussed. For example, "Use the maximum possible number of modalities for each element" could be one of them. More elaborated categories of criteria are the following:

- Technical Criteria: Examples for this type of criterion are, "Use the modalities which can represent the output in the fastest way" or "Use the modalities whose corresponding devices can execute the output with the lowest power consumption". This kind of data must be queried from the concrete devices at runtime.

- Context-Aware Criteria: The dialog history can be used to receive an output which is adjusted to the current conversational situation. For example, if the robot points to an object when it is mentioned first, it is not necessary to point to it again in the following sentences. Instead, it is sufficient to reference the object via speech. Moreover, by using information retrieved from the user model, the output can be adapted to the current user. For example, impairments, language skills or preferred output modalities can be taken into account. Furthermore, information about the environment, like noise level or lighting conditions, can be considered.

- Human-Likeness Criteria: This category contains all criteria which have the aim to let the robot act in a more human-like way while outputting the information. For example, a general criterion would be to look at people rather than to point to them. Besides, looking at the person to talk to, not using the Pointing Modality for each word of the sentence or looking at things while pointing at them might also lead to a more human-like behavior. Furthermore, speech output is the foundation of the conversation while the other modalities clarify and emphasize certain things. Additionally, considering Context-Aware Criteria, like the dialog history and the user model mentioned above, also improves human-like behavior.

Some criteria from each of the mentioned categories are already implemented. It is up to the user of the framework and his or her presentation goals to choose which criteria to use. New criteria, particularly adapted for the use of certain modalities, can be added in an easy manner. How well the implemented criteria, especially the Human-Likeness Criteria, perform or how to further improve them, will be evaluated in this thesis.

DISCUSSION

Several Multimodal Dialog Systems have been developed in the last twenty years. Such systems are highly complex and, in most cases, their main focus is not on the Fission Module. Because of this, a lot of these systems hardcode the output presentation or adapt it to their used domain.
Therefore, in this thesis, a framework for Multimodal Fission is developed. The MMF framework, which is implemented in Java, is extendable and domain-independent. Its focus is on Human-Robot Interaction.
The framework only requires a predicate using the semantics of the corresponding predicate in the artificial language Lojban as an input from a potential Dialog Manager. This will facilitate the task of connecting the Fission framework to any arbitrary Dialog Manager. For adding new predicates to the framework, only the predicate structure of the corresponding predicate in Lojban needs to be looked up in the previously referenced English-Lojban Dictionary. That means, the user of the framework does not need to define a suitable predicate structure by his or her own and the used predicates all follow the same principles, which have been developed for Lojban by a large number of contributors over four decades.
Moreover, the framework can be extended by adding new modalities. Two types of modalities, namely Structure-Forming and Object-Referencing Modalities, have been implemented. When adding a modality which does not fit into these two categories, new categories might be added, which need to define the tasks of its associated modalities. Furthermore, the set of used modalities can be modified. However, a Structure-Forming Modality is required, because it defines the different output parts and the order in which they will appear in the output.
New devices can also be added by the user of the framework. To do this, he or she must define how the execution of an output element on a concrete device is realized.
Apart from that, the framework can be extended by new planning criteria for the modality and device selection process. The new criteria can be added in the form of a new score calculator, which will then be passed to OptaPlanner. Alternatively, an existing score calculator can be extended by new planning criteria. A weighted function can be used to determine the importance of each criterion. Providing different criteria enables the framework to choose the most suitable modalities and devices for the current user in the current environment and in the current conversational situation.
Information about the domain is encoded in the world model. The domain independence is realized by the fact that the world model is an input of the framework, which can be updated or exchanged if necessary. The world model can be provided in different formats, which increases flexibility. For now, ontologies authored in OWL and the document-orientated database MongoDB are supported. Document databases like MongoDB use JSON documents in order to store records. Since JSON-Objects are internally used to represent the world model, it is very easy to extract the information from the database and store it in the internal representation. Furthermore, in contrast to simple key-value stores, document-orientated databases are able to store nested key-value pairs.
The implemented modalities were chosen with a focus on Human-Robot Interaction. Besides, since they are Object-Referencing Modalities, Multimodal References can be realized. The Unique Identifier Algorithm has been designed to resolve references via speech. The algorithm extracts properties which enable a differentiation between objects of the same type. Any delay on the availability of the world model containing these properties, as well as incomplete description of the objects, is not yet considered. However, the algorithm might be improved by taking the other modalities into account. Instead of using a long and complicated description for referencing, it might also be possible to use a property (not necessarily unique) in combination with a pointing action so that the referred object can be uniquely identified. This approach could also be used if a differentiation is important, but a differentiating property does not exist.
Furthermore, the MMF framework will be compared to the Fission component used in a multimodal dialog framework called Ontology-based Dialogue Platform (ODP) [5]. This framework enables the creation of multimodal dialog applications for new domains by using a model-driven development. It was presented in the scope of THESEUS[4], a research program for the development of an internet-based knowledge infrastructure.

FUTURE WORK

After having implemented a first version of the MMF framework, the framework will be extended. Currently, the predicates used as input can only encode statements. A future version will also support questions and commands.
Furthermore, the timing information of the speech synthesizer could be used to enable a better output coordination. For now, the output plan can be seen as a sequence of parallel actions. This means that information presented by different modalities can either be executed in parallel or after each other. The timing information of the speech synthesizer can be used to create a preliminary output schedule. Furthermore, after selecting

[4]THESEUS: http://foerderprogramm-theseus.dfki.de/

the concrete modalities and devices, the execution duration for each part of the output can be estimated. This information can then be used to create a final schedule which defines for each part of the output a discrete starting time. This enables

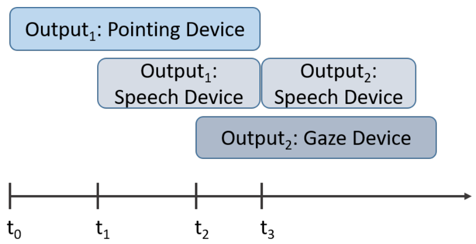


Figure 5. Example of an interleaved Output

an interleaving of parts of the output. An example for an interleaved output can be seen in Figure 5.
In addition, the speech output will be improved by exchanging the current Natural Language Generation tool.
Another extension might be to supplement the three implemented modalities with a fourth one. It is yet to be decided whether to use a modality which can present images and graphics or one which can perform gestures. Both could be helpful to improve the interaction between human and robot.
Apart from that, the performance of the already implemented planning criteria for the modality selection will be evaluated and new planning criteria will be added. For example, a more sophisticated dialog history might be used.
The framework currently supports Multimodal References, but also *Deictic References* will be considered. The term *Deixis*[5] describes words which cannot be entirely understood without information about the context. *Gestural deixis* refers to deictic expressions that require some kind of audio-visual information. Pointing to an object and referring to it as "this" or "that" is an example for gestural deixis. This kind of reference will be supported.
After extending the framework, a user study will be conducted to evaluate the framework's multimodal output. In this study, the multimodal output produced by the framework will be compared to the multimodality which is naturally used by humans when interacting with each other. Additionally, it can be evaluated how human-like the robot's behavior is while performing the output. Furthermore, it will be evaluated if a better understanding of the presented information is enabled if a combination of modalities is used, compared to a presentation using only speech output. It will be investigated whether the multimodal output always enables a better understanding or whether a combined modality usage might lead to confusion in some situations. These results will then be compared to existing studies about multimodal output.

CONCLUSION

The aim of this thesis is to create an extendable and domain-independent Multimodal Fission Framework with a focus on Human-Robot Interaction. The motivation behind the use of multimodal output is to enable a more human-like conversation. Some contributions in the area of Fission, in particular the theoretical categorization of the tasks in Fission, have been introduced. Furthermore, an overview of the MMF framework has been given. The input to the framework consists of a world model and a predicate and it outputs triples consisting of the used modality, device and output element. New predicates can be added in a straightforward manner, due to the fact that they are based on Lojban's predicate structure. Moreover, the new terms *Structure-Forming* and *Object-Referencing* Modality have been introduced to categorize modalities. Besides, to enable reference resolution via speech, an algorithm called *Unique Identifier Algorithm* has been designed. Finally, the modality and device selection process, based on different planning criteria, has been discussed. Several criteria are already offered, but new ones can be added easily.

[5]Deixis: http://www-01.sil.org/linguistics/GlossaryOfLinguisticTerms/WhatIsDeixis.htm

REFERENCES

1. David Costa and Carlos Duarte. 2011. Adapting multimodal fission to user's abilities. In *International Conference on Universal Access in Human-Computer Interaction*. Springer, 347–356.

2. Geoffrey De Smet et al. *OptaPlanner User Guide*. Red Hat and the community. http://www.optaplanner.org OptaPlanner is an open source constraint satisfaction solver in Java.

3. Mary Ellen Foster. 2002. State of the art review: Multimodal fission. *COMIC project Deliverable* 6, 09 (2002).

4. Mary Ellen Foster and Michael White. 2005. Assessing the impact of adaptive generation in the COMIC multimodal dialogue system. In *Proceedings of the IJCAI 2005 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*.

5. Robert Neßelrath and Daniel Porta. 2011. Rapid development of multimodal dialogue applications with semantic models. *Knowledge and Reasoning in Practical Dialogue Systems* (2011), 37.

6. Cyril Rousseau, Yacine Bellik, Frédéric Vernier, and Didier Bazalgette. 2006. A framework for the intelligent multimodal presentation of information. *Signal Processing* 86, 12 (2006), 3696–3713.

7. Wolfgang Wahlster. 2003. Smartkom: Symmetric multimodality in an adaptive and reusable dialogue shell. In *Proceedings of the human computer interaction status conference*, Vol. 3. Berlin (Germany): DLR, 47–62.

8. Wolfgang Wahlster, Elisabeth André, Som Bandyopadhyay, Winfried Graf, and Thomas Rist. 1992. WIP: The coordinated generation of multimodal presentations from a common representation. In *Communication from an Artificial Intelligence Perspective*. Springer, 121–143.

D Technical Reports

Ontology-based
Content planner Grammars for PAL

Stefania Racioppa, DFKI GmbH

February 2, 2017

Abstract

This technical report describes the work on the Content Planner [2] grammars in PAL . In the past year, the main focus lied on the alignment of the Speech act definitions to the frames and instances defined in the PAL ontology. Besides this, the Content Planner and OpenCCG grammars were constantly improved in terms of coverage and parsing precision.

1 Overview

The Content planner [2] is responsible in PAL for the modeling of the verbal output. This output is currently defined for English and Dutch in a large range of activity-related speech acts in *string* form. Every output may consist of a single or more sentences, which can be concatenated or used alternatively.

Due to the strong inflection and agreement rules in Italian, the speech acts output for this language is expressed by *logical forms* instead, which are based on the semantic output of the Italian OpenCCG [5] grammar. As features can be parameterized more easily than strings, i.e. just using simple variables, substituting the output strings with logical forms makes the rule maintenance more comfortable.

Using the OpenCCG surface realizer, we then can build from the defined logical forms the surface form of the required utterance, which again is the input for the TTS system [4].

2 Grammar upgrade

The Content planner and OpenCCG grammars were continuously improved to clear realization errors and get rid of some cases of overgeneration. We detected and solved duplicate or conflicted rules and completed the coverage of the speech acts required by the dialogue manager. Some deprecated parameters were reactivated in the current version (e.g. `PerformanceEval`).

In preparation for the implementation of the ontology-based speech act definition, we made sure that the relevant speech acts are defined in the same way in all project languages (EN, NL, and IT), and the rules generate a similar output, or a similar number of utterance variants. This alignment is important not only for the coverage consistency, but also to ensure that the language-independent *mapping layer* we realized for this task (see 3.1) works properly.

```
:dvp ^ <SpeechAct>greeting ^ <Context>(<RobotName>#robot ^ <Encounter>first)
->
###x = concatenate(random("hallo , mijn naam is ", "hoi , ik ben ", "hallo , ik ben ",
                   "hoi, mijn naam is "), #robot),
# ^ :canned ^ <stringOutput>###x ^ <SpeechModus>indicative.
```

```
:dvp ^ <SpeechAct>greeting ^ <Context>(<RobotName>#robot ^ <Encounter>first)
->
###x = concatenate(random("hello , my name is ", "hi , I am ", "hello , I am ",
                   "hi , my name is "), #robot),
# ^ :canned ^ <stringOutput>###x ^ <SpeechModus>indicative.
```

3 Ontology-based speech act definition

The ontology developed at DFKI for the PAL project integrates among others ideas from the DIT++ taxonomy of dialogue acts (see `http://dit.uvt.nl`), which includes some typical arguments used in the Content planner speech act definition (e.g. dialogue act *sender* or *addressee*), with project-specific knowledge items (e.g. *Actor*, *Activity*). [3]

The dialogue act models covered by the PAL ontology already offer a *comprehensive, application-independent system for the analysis of human and human-machine dialogue* [1]. For this reason, we decided to use the ontology frames and instances to normalize the Content planner speech acts definitions.

```
greeting(Encounter=first) -> InitialGreeting(Meeting)
accept(role, Asker=robot) -> Accept(AssigningRole, agent=robot)
apologize(question) -> Apology(Asking)
```

As the defined speech acts are strictly related to the activities planned in the dialogue manager, we had to integrate the existing taxonomy with domain-specific definitions in some cases (e.g. `BeingCorrect`, or *manner*: `Switch`, and `Repeat` vs. `NoRepeat`).

3.1 Implementation

A first attempt to modify the Content planner grammars *directly* failed because of the complex interdependency between the rules [4]. This approach would also have made it necessary to modify the rule sets of all project languages separately, increasing the risk of inconsistency and errors. For this reason, we left the grammars unchanged at the end, and implemented a *common* rule set as a kind of *mapping layer* instead, which transforms the ontology-based definitions to the ones used in the single grammars.

The advantage of a language-independent, top-level mapping stage is evident: changes in the ontology or grammars can be easily implemented in a central rule set in the future, and will apply automatically to all project languages.

In the most cases, we could easily map the ontology dialogue act items (e.g. `Inform`, `Apology`) to the main PAL speech acts (e.g. `provide`, `apologize`) and their arguments.

```
<SpeechAct>(#s ^ Apology) ^ <Content>#c: -> #s = apologize.
<SpeechAct>(#s ^ CheckQuestion) ^ <Content>#c: -> #s = confirm.
<SpeechAct>(#s ^ Inform) ^ <Content>#c: -> #s = provide.
<SpeechAct>(#s ^ Request) ^ <Content>#c: -> #s = request.

<Content>#c:
{
  ^ AssigningRole -> #c ^ <About>role, # ! <__PROP>.
  ^ BeingCorrect -> #c ^ <About>correctness, # ! <__PROP>.
  ^ BeingSuccessful -> #c ^ <About>success, # ! <__PROP>.
}
```

This one-to-one conversion wasn't possible in some cases, because of the granularity differences between ontology and grammar definitions. The corresponding *mapping* rules required the definition of one or more additional features in the LHS or RHS, depending whether the Content planner or the ontology provides the most detailed taxonomy.

```
:InitialGreeting ^ Meeting
->
# ^ :greeting ^ <Context>(<Encounter>first),
# ! <__PROP>.

:ReturnGreeting ^ Meeting
->
# ^ :greeting ^ <Context>(<Encounter>notfirst),
# ! <__PROP>.

<SpeechAct>(#s ^ Confirm) ^ <Content>#c: -> #s = acknowledgement, #c ^ <Value>yes.
<SpeechAct>(#s ^ Disconfirm) ^ <Content>#c: -> #s = acknowledgement, #c ^ <Value>no.

<Content>#c:
{
  ^ Playing ->  #c ^ <About>play, # ! <__PROP>.
  ^ Playing ^ <manner>Continue -> #c ^ <About>playContinue, # ! <manner>, # ! <__PROP>.
  ^ Playing ^ <manner>Switch -> #c ^ <About>switch, # ! <manner>, # ! <__PROP>.
}
```

References

[1] Harry Bunt. *DIT++ Taxonomy of Dialogue Acts*. `http://dit.uvt.nl`, 2010.

[2] Bernd Kiefer. *Content Planner Module for Talking Robots.*, 2014.

[3] Hans-Ulrich Krieger. *Ontologies and Reasoning Architecture for PAL.*, 2015.

[4] Stefania Racioppa. *OpenCCG and LF Grammars for PAL.*, 2015.

[5] Mark Steedman. *The syntactic Process*. The MIT Press, Cambridge Mass., 2000.